

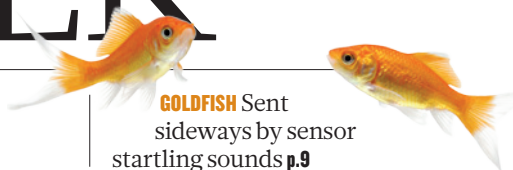
THIS WEEK

EDITORIALS

FUTURES 400 and counting for *Nature's* science-fiction column **p.6**

WORLD VIEW The problem with consensus reports is the consensus **p.7**

GOLDFISH Sent sideways by sensor startling sounds **p.9**



High-interest clones

Research into cloned human cells has left the spectre of past scientific fraud behind. But reaction to the earlier work still holds worthwhile lessons.

When, in 2004, Woo Suk Hwang claimed to have produced a stem-cell line derived from an embryonic human clone, his research, done at Seoul University, sparked intense interest and hype. Even though Hwang's work later proved to be fraudulent, all advances in the field risk being measured against it. At the same time, researchers seek to distance themselves from the episode to the extent that its ethical implications for current work are rarely discussed.

This week, scientists have come the closest of any so far in emulating Hwang's claimed results: on page 70, researchers from the New York Stem Cell Foundation Laboratory report using cloning technology to reprogram human DNA taken from an adult and create embryonic stem cells. But they do not use the term cloning to describe their results. That is one of many contrasts between the research landscape now and in 2004.

Hwang's claims received worldwide attention. Patient groups jumped for joy; scientists around the world used the results to gather more funds for stem-cell research; and bioethicists emerged to justify or condemn the work. Reaction this week is likely to be more muted.

Discussion of the ethical concerns raised by such work have calmed, and the research group behind the latest study dealt with one of the most divisive issues — the retrieval of human eggs from donors — in a transparent and considered way. Hwang, by contrast, had procured eggs unethically and illegally, a problem first brought to public attention in *Nature* (see *Nature* **429**, 3; 2004). Whereas discussion of Hwang's results featured the phrase 'therapeutic cloning' and so invited (sometimes wilful) confusion with reproductive cloning and the spectre of technology misuse, the latest paper refers only to the reprogramming of cells to a pluripotent state. A final issue — that embryos are destroyed in the process of the research — does still apply.

The ultimate goal of such research is to create patient-specific stem cells for drug screening and the growth of genetically identical tissue for transplantation. Yet cloning, whether called that or not, is no longer the only means to this end, as it seemed in Hwang's time. Induced pluripotent stem (iPS) cells, first developed in 2006, now offer the same promise without the need for egg recruitment or embryo destruction: they are produced from adult cells by introducing a few genetic factors to the cell rather than using an entire egg. When therapeutic-cloning studies stalled on an egg shortage, iPS cell frenzy filled the gap. Competition between the approaches is fierce, and the authors of the current study point out the many weaknesses of iPS cells to bolster their own work. But their approach, too, has a long way to go.

The biggest reason that the results won't generate Hwang-like headlines is that they do not go as far. Hwang claimed to have created a cloned human embryo with the same 46 chromosomes as its parent, in a very similar way to how scientists have produced living cloned mammals. Hwang's embryo would have been viable, generating huge ethical debate. His claimed results were so advanced that in 2005, Hwang was applying to start clinical trials.

The cells presented this week have an 23 extra chromosomes from the egg. Hwang, like most researchers in the field, removed this DNA and used the egg merely to drive reprogramming; it didn't work. The latest study left the egg DNA in, and says that some element of it is essential.

The cells derived from this 'triploid' embryo show many of the functions of normal cells, but such embryos are not viable and it is

not yet clear how triploid cells would mimic the behaviour of cells in tissue. No one will be calling them clinically relevant any time soon.

Still, iPS cell work is on the defensive, and this study provides proof that human somatic cells can be reprogrammed.

Now, researchers have to prove that the work is a step towards a biomedically useful stem-cell line. The authors are confident that they can produce a stem-cell line from

a 'normal' diploid cloned embryo, as Hwang claimed to do. They will have to work out what it is in the egg's genetic material that is necessary for the reprogramming.

The latest achievement points in the same direction as Hwang's claims. If researchers were to find the magic element in the egg, not only would there again be excitement, but the old ethical issues would resurface. Hype around potential procedures would increase the market for eggs, which is perhaps hard to justify. The embryos would be viable, no doubt again producing fears of self-cloning dictators. (For that reason, this might be a good time for the United Nations to hammer out cloning regulations or restrictions, which have been hamstrung by political and religious debate.) And desperate patients would find doctors ready to give them unproven and unsafe embryonic-stem-cell treatments.

The results might look mundane. But the potential for reasoned excitement and irrational hype remain. ■

The games begin

Frustrations of the newest European member states will shape debate over research funding.

With some €80 billion (US\$105 billion) to distribute, the next European research funding programme will have one of the world's most generous science budgets. The European Commission has promised radical change to the programme, called Horizon 2020, and researchers, politicians and commentators have been waiting to see the results. This week, *Nature* reveals the programme's new look.

A leaked draft of the commission's plans for Horizon 2020,

discussed on page 16, reveals the commission's admirable and much-needed attempt to make applying for funding and participating in the programme a lot easier for researchers.

It is true that the commission has attempted to streamline past programmes, notably the current Seventh Framework Programme, which runs until 2013. But wider and deeper change is needed — particularly, as the commission now suggests, to harmonize the criteria for evaluating research proposals and judging what counts as eligible research costs across the different sections of the programme. This means that researchers will need to learn only one set of rules, whether they apply for funding for collaborative research projects that tackle key societal challenges such as energy efficiency, or for grants from, for example, the Budapest-based European Institute of Innovation and Technology.

The proposals also suggest the provision of better guidance for researchers who must fill out time sheets to satisfy the commission's demand for financial accountability once research projects are complete. Significantly, the plans abolish time recording for researchers who work full-time on one project, such as those with grants from the European Research Council, which funds frontier research across Europe. This comes as a welcome move, as there is little point in having all that money available if the rules of play are so complicated and time-consuming that they discourage researchers from applying. The cream of Europe's science crop have, in the past, turned their backs on the funding programme to compete for other, less-bureaucratic funding streams, threatening to push the Framework programme towards mediocrity.

There is no guarantee that the revisions go far enough to halt this trend. And as the finer details of the programme are hammered out, including developing the annual calls for proposals, the commission should allow researchers more freedom to draw up research proposals, rather than continuing to prescribe the precise projects that it wants to fund. It is scientists, not Brussels bureaucrats, who are best placed to know what is new and interesting.

Missing from the leaked proposals is a clear solution to the tension building among the 12 newest European member states — known as the EU 12, including Poland and Romania — which feel excluded by the drive to fund excellent science. With weaker national science

and technology systems, researchers in these countries often lose out to their counterparts in the scientifically stronger nations, who are able to write better grant proposals. Researchers in the EU 12 complain that young research talent is not being given the support or the opportunity to show its potential. They are not alone in their concerns over the uneven geographical distribution of the programme's funds.

"Nature applauds the commission's hard-fought efforts to prioritize excellence as a key funding criterion."

Members of the European Parliament's industry, research and energy committee said in a report on 31 August that they find it unacceptable that the lion's share of research funds goes to the richer member states.

Traditionally, the commission allocates support for national capacity-building through a separate funding stream available in the European Union's budget, called structural funds. The commission encourages their use for research purposes in the newer member states. But it has had mixed results, in part because governments prefer to use the funds for improvements that their voters can see and use, such as new roads. It is not enough for the commission to claim that structural funds can help to put newer member states on the path towards excellent science, as it does in the Horizon 2020 draft. Rather, it must propose concrete initiatives and reforms that encourage those governments to use these funds for research. A key starting point could be to cut the red tape around the use of structural funds, which is even more difficult to navigate than the research Framework programme.

Nature applauds the commission's hard-fought efforts to prioritize excellence as a key funding criterion — specifically, its plan to devote one-third of the programme to excellent research. This focus will be ever more important if Europe is to compete on the global research stage. Nevertheless, the frustrations of the EU 12 countries need to be addressed, not least because they, along with members of the European Parliament, could delay agreement on the programme plans. This issue is likely to dominate much of the debate on the shape of European funding over the next 18 months. So, let the games begin. ■

Back to the Futures

As Nature's science-fiction column reaches a milestone, we recall some of the highlights.

This week sees the 400th science-fiction story published in *Nature* journals under the 'Futures' banner. The number 400 is, of course, only significant to those of us with ten digits. It's more impressive in binary (110010000), although nothing special in Hex (190th), and the Octalonacci of the Octillian system (our keenest readers) will mark it as their 620th.

The number, however presented, includes all the stories we have published in *Nature* — on, off, simultaneously or instead of — since Arthur C. Clarke's inaugural salvo on 4 November 1999, as well as those featured in the completely separate time-stream of *Nature Physics*, a few parsecs away.

Looking back at the Futures, as they say, we find that the column, while barely noticed by many, sitting as it does at the back of each printed issue (although free to all online), is a guilty pleasure for the discerning few. The anthology, *Futures from Nature*, was given a starred review by *Publishers Weekly*, and, in 2005, the column won *Nature* the accolade of 'Best Science Fiction Publisher' from the European Science Fiction Society. (We ignore those wags who say that everything that *Nature* publishes is science fiction.)

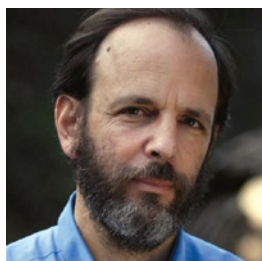
Among the canon of stories published in Futures are missives from

superluminaries of the genre: Michael Moorcock, Frederik Pohl and Ursula Le Guin. (Had Isaac Asimov been alive, he'd probably have written the lot.) But there have also been tales from other established writers, perhaps less well known to *Nature* readers, and many more from scientists — and others — trying out fiction for the very first time. We've had lesbian robots from a senior citizen in Alaska, the shade of Michael Jackson from a virologist in Singapore, the problems of copyrighting dreams from a software consultant in India, and intergalactic country music by a student from Malaysia. Futures was also the venue for the first story ever sold by high-school student Shelly Li of Omaha, Nebraska — who is now just about to publish her first novel.

You, too, can join the throng by following the exploits of Futures on Facebook (go.nature.com/mtoodm) or by sending your story (850–950 words) to futures@nature.com. But beware, Futures has become a victim of its own success — like trying to nail jelly to the ceiling, it is now almost as hard to get a story into Futures as to have a research paper published in *Nature*.

Futures, like radio signals from distant suns, will surely come and go. But as the man said (the 'man', depending on which web page you read, being Yogi Berra, Niels Bohr, Woody Allen or, who knows, Donald Rumsfeld), prediction is very difficult, especially about the future. As such, we intend to keep Futures until we (or you) get bored of it, or until Earth is struck by an asteroid, whichever comes first. The first seems unlikely. As for the second, we shall no doubt have other, more pressing concerns. Here's to the next 110010000. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



The voice of science: let's agree to disagree

Consensus reports are the bedrock of science-based policy-making. But disagreement and arguments are more useful, says Daniel Sarewitz.

When scientists wish to speak with one voice, they typically do so in a most unscientific way: the consensus report. The idea is to condense the knowledge of many experts into a single point of view that can settle disputes and aid policy-making. But the process of achieving such a consensus often acts against these goals, and can undermine the very authority it seeks to project.

My most recent engagement with this form of penance is marked this week with the release of *Geoengineering: A National Strategic Plan for Research on Climate Remediation*. Sponsored by the Bipartisan Policy Center in Washington DC, the report reflects more than a year of discussion between 18 experts from a diverse range of fields and organizations. It sets out, I think, many valuable principles and recommendations.

The discussions that craft expert consensus, however, have more in common with politics than science. And I don't think I give too much away by revealing that one of the battles in our panel was over the term geoengineering itself.

This struggle is obvious in the report's title, which begins with 'geoengineering' and ends with the redundant term 'climate remediation'. Why? Some of the committee felt that 'geoengineering' was too imprecise; some thought it too controversial; others argued that it was already commonly used, and that a new term would create confusion.

I didn't have a problem with 'geoengineering', but for others it was a do-or-die issue. I yielded on that point (and several others) to gain political capital to secure issues that had a higher priority for me. Thus, disagreements between panelists are settled not with the 'right' answer, but by achieving a political balance across many of the issues discussed.

This political essence of consensus leads to other difficulties. Ask a panel to address broad questions — future directions for a field, say, or ways to improve a government programme — and the recommendations that come back are typically bland and predictable. New and controversial ideas are inherently difficult for experts to agree on. In the absence of consensus, the default position is simply to call for more research — the one recommendation that most scientists can get behind.

Sometimes, expert panels are asked to find consensus on narrow technical questions at the heart of public controversies. The hope is that a unified scientific voice will resolve the dispute, but it rarely works out that way. In 2000, the US National Academies assembled climate experts to resolve discrepancies in surface and satellite climate temperature records, as if this would help to settle the political debate. A decade on, it is clear that the goal was not met.

And in 2009, at the height of the US debate on health-care reform, the US Preventive Services Task Force released a consensus report on the risks and benefits of mammograms. Rather than

clarifying anything, the key recommendation — that mammograms were being overutilized — became instant ammunition for reform opponents, who viewed it as a threat to patient autonomy.

The fuss over mistakes in the 2007 reports by the Intergovernmental Panel on Climate Change highlights a related problem: a claim of scientific consensus creates a public expectation of infallibility that, if undermined, can erode public confidence. And when expert consensus changes, as it has on health issues from the safety of hormone replacement therapy to nutritional standards, public trust in expert advice is also undermined.

The very idea that science best expresses its authority through consensus statements is at odds with a vibrant scientific enterprise. Consensus is for textbooks; real science depends for its progress on continual challenges to the current state of always-imperfect knowledge. Science would provide better value to politics if it articulated the broadest set of plausible interpretations, options and perspectives, imagined by the best experts, rather than forcing convergence to an allegedly unified voice.

Yet, as anyone who has served on a consensus committee knows, much of what is most interesting about a subject gets left out of the final report. For months, our geoengineering group argued about almost every issue conceivably related to establishing a research programme. Many ideas failed to make the report — not because they were wrong or unimportant, but because they didn't attract a political constituency in the group that was strong enough to keep them in. The commitment to consensus therefore comes at a high price: the elimination of proposals and alternatives that might be valuable for decision-makers dealing with complex problems.

Some consensus reports do include dissenting views, but these are usually relegated to a section at the back of the report, as if regretfully announcing the marginalized views of one or two malcontents. Science might instead borrow a lesson from the legal system. When the US Supreme Court issues a split decision, it presents dissenting opinions with as much force and rigour as the majority position. Judges vote openly and sign their opinions, so it is clear who believes what, and why — a transparency absent from expert consensus documents. Unlike a pallid consensus, a vigorous disagreement between experts would provide decision-makers with well-reasoned alternatives that inform and enrich discussions as a controversy evolves, keeping ideas in play and options open. That is something on which we should all agree. ■

Daniel Sarewitz is co-director of the Consortium for Science, Policy and Outcomes at Arizona State University, and is based in Washington DC. e-mail: daniel.sarewitz@gmail.com

REAL SCIENCE
DEPENDS FOR ITS
PROGRESS
ON CONTINUAL
CHALLENGES TO THE
CURRENT STATE OF
ALWAYS-IMPERFECT
KNOWLEDGE.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/9ohmyy

SEVEN DAYS

The news in brief

POLICY

NIH-budget duel

A bill released on 29 September by the spending committee of the US House of Representatives would boost the budget of the National Institutes of Health (NIH) by US\$1 billion, or 3.3%, in 2012. That puts it in direct conflict with a Senate bill cutting the agency's budget by \$190 million, to \$30.5 billion, next year. The two bills must be resolved in the coming weeks. The Senate bill also establishes and funds a proposed translational medicine centre at the NIH, but the House bill does not mention it.

Energy priorities

The US Department of Energy released its inaugural Quadrennial Technology Review on 27 September, laying out a multi-year agenda that sets priorities in six areas, including vehicle efficiency, alternative hydrocarbon fuels and cleaner electricity. It is modelled on the Quadrennial Defense Review, an analysis that sets the tone and direction of US defence policy. The first energy-technology review made no radical recommendations, mostly keeping to the agency's current course. See go.nature.com/uyabvv for more.

RESEARCH

SPICE on ice

A UK experiment to test climate-engineering technology by spraying water from a balloon 1 kilometre above Earth was put on hold this week — owing partly to protests from environmentalists. The Stratospheric Particle Injection for Climate Engineering (SPICE) project aims to trial technology for spraying

sulphate aerosols at heights of up to 25 kilometres, to cool Earth by reflecting sunlight. Protesters said that the test would violate a decision by the Convention on Biological Diversity not to undertake large-scale geoengineering experiments. SPICE scientists say that the halt followed a consultation which raised concerns that there had not been enough engagement with environmental groups. See go.nature.com/hmuljg for more.

Jackson expands

The Jackson Laboratory, a medical research centre based in Bar Harbor, Maine, is hoping to set up a major satellite facility for personal genomics research at the University of Connecticut

Health Center in Farmington. The lab had previously tried for more than a year to site its facility in Florida, but had to pull out in June (see *Nature* **474**, 133; 2011) after Florida politicians said that the state could not afford to invest in it. But Connecticut is prepared to contribute US\$291 million towards the \$1.1-billion lab, governor Dannel Malloy announced on 30 September. The deal depends on agreement from the state's legislature.

Mouse phenome

An ambitious effort to work out the function of all of the approximately 21,000 protein-coding genes in the mouse genome has the funds it requires for the initial phase of its mission. After a meeting

the top quark was found in 1995 — and it spent its final years restricting the possible mass range of the Higgs boson. Scientists are still analysing those data, and Fermilab is shifting to smaller-scale experiments (see *Nature* **477**, 379; 2011).

Farewell to the Tevatron

After more than 25 years of colliding particles, the massive Tevatron accelerator at Fermilab in Batavia, Illinois, was turned off for good on 30 September. The collider helped to confirm the standard model of physics — it was where



R. HAHN/FERMILAB

in Washington DC last week, the International Mouse Phenotyping Consortium, which was launched last year (see *Nature* **465**, 410; 2010), announced that research agencies in eight countries will analyse 5,000 mouse genes by 2016, with part of the work being financed by the US National Institutes of Health. The work involves inactivating a particular gene and then investigating how the mouse's characteristics change.

Global eco-network

Plans for a global network to monitor agricultural areas came a step closer to being realized last week, as scientists, philanthropic organizations and big businesses met at Columbia University in New York to discuss the idea.

CHINA/PHOTO PRESS/GETTY IMAGES

Researchers described pilot sites in Africa that collect data on soils, nutrients and land cover, but also track changing agricultural practices and socioeconomic trends. Jeffrey Sachs, director of Columbia's Earth Institute, said he hoped that — with industrial funding — the network might grow to some 500 sites in two or three years. See go.nature.com/yu5qdt for more.

Graphene cash

The UK government is investing £50 million (US\$78 million) to create a research and technology hub focusing on graphene, the one-atom-thick sheets of carbon for which physicists Andre Geim and Konstantin Novoselov, based at the University of Manchester, UK, won the 2010 Nobel Prize in Physics. British scientists — still smarting from broad funding cuts — welcomed the 3 October announcement, which also included £145 million to build infrastructure for high-performance computing. See go.nature.com/7pzaio for more.

China space lab

Tiangong-1, a test module for China's space station, was launched on 29 September (pictured). The 10.4-metre cylinder is orbiting alone and unmanned for now, but in a few weeks' time China will



try to dock it in orbit with an unmanned spacecraft. Two missions carrying Chinese astronauts will follow in 2012, with more test modules over the subsequent three years. If all goes to plan, China will launch further modules to be assembled into a space station by 2020 (see *Nature* 473, 14–15; 2011).

HIV prevention

Hopes of preventing HIV by giving drugs to uninfected women were dented last week when part of a large clinical trial testing the idea was halted. The Microbicide Trials Network said that it will no longer give out tenofovir antiretroviral tablets in the VOICE study — involving 5,029 women across South Africa, Zimbabwe and Uganda — after an independent monitoring board found that, on the basis of results so far,

it wasn't possible to show that the tablets were better than placebo. Other arms of the trial, involving a gel and another antiretroviral tablet, continue. Previous trials have yielded contradictory results on whether the practice works for women, although it has been shown to reduce HIV risk for men.

BUSINESS

Amazon dam halted

Construction of the Belo Monte Dam — a massive 11.2-gigawatt hydroelectric plant on a tributary of the Amazon in Brazil — has for the second time this year been put on hold by a federal judge, who ruled that it could damage fish stocks. The dam, on the Xingu River in the state of Pará, is opposed by environmental campaigners and indigenous people, but the government is strongly behind it, and the environment agency IBAMA has already granted a construction licence. The Norte Energia consortium of companies building the dam will appeal against the ruling.

PEOPLE

US science medals

Cloning pioneer Rudolf Jaenisch, a biologist at the Whitehead Institute in Cambridge, Massachusetts, is one of seven researchers to

COMING UP

9–12 OCTOBER

The Geological Society of America meets in Minneapolis, Minnesota.

go.nature.com/5vvlxp

10–21 OCTOBER

Delegates from 194 countries will gather in Changwon, South Korea, to discuss progress and targets to tackle desertification at the 10th meeting of the United Nations Convention to Combat Desertification.

go.nature.com/8zq4jl

11–15 OCTOBER

The 12th International Congress of Human Genetics takes place in Montreal, Canada.

www.ichg2011.org

be awarded the US National Medal of Science this year. The White House announced the list of recipients — along with five awardees of the National Medal of Technology and Innovation — on 27 September. The medals are the highest honours that the United States bestows on its scientists and engineers. See go.nature.com/j78ggo for more.

Nobel prizes

This year's Nobel Prize in Physiology or Medicine went to Bruce Beutler, Jules Hoffman and Ralph Steinman, for their work on the immune system. The physics prize was won by Saul Perlmutter, Brian Schmidt and Adam Riess, for discovering the accelerating expansion of the Universe by observing distant supernovae. See pages 13 and 14 for more. *Nature* went to press before the chemistry prize was awarded, but full details will be available at go.nature.com/uio77d.

➔ NATURE.COM

For daily news updates see:

www.nature.com/news

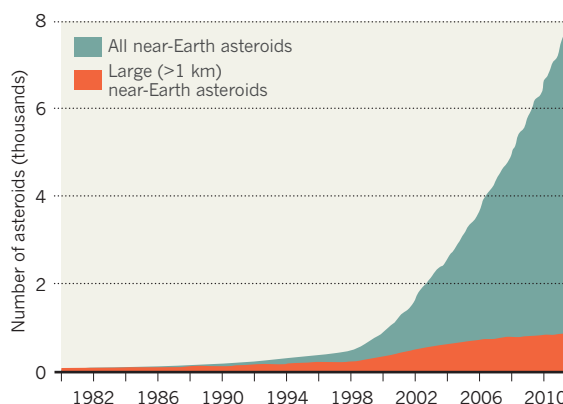
SOURCE: NASA/JPL

TREND WATCH

NASA said on 29 September that it had spotted 911 of the estimated 981 near-Earth objects larger than 1 kilometre across, using infrared data to recalculate the relationship between asteroid reflectivity and size. The agency has now met the US Congress's 1998 mandate to find more than 90% of 'killer' asteroids. A 2005 mandate extended that to asteroids down to 140 metres across; so far, NASA thinks it has found 35% of the estimated 13,200 such objects.

NASA REACHES ASTEROID-SPOTTING GOAL

The agency has found 90% of the asteroids larger than 1 kilometre near Earth, plus thousands that are as small as a few metres across.

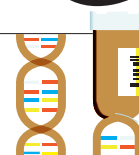


NEWS IN FOCUS

PUBLIC HEALTH Brain disorders: staggering costs, modest investments **p.15**

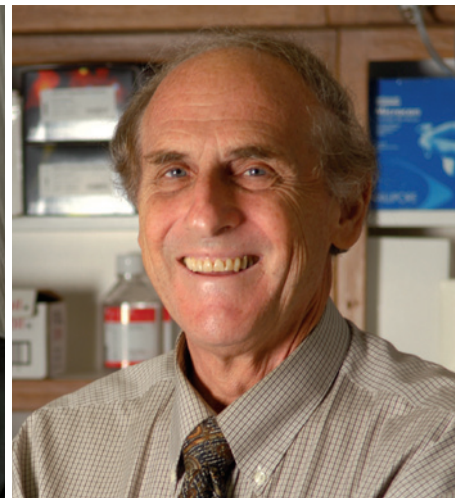
EUROPE Mega-funding programme aims for minimum hassle **p.16**

ETHICS Should subjects be told if their DNA reveals a nasty surprise? **p.17**



GENOMICS Full sequencing becomes a clinical tool **p.22**

(FROM LEFT) C. BARRIA/REUTERS; SCRIPPS RES. INST.; ROCKEFELLER UNIV.



Immunity service: (from left) Jules Hoffmann, Bruce Beutler and Ralph Steinman share this year's Nobel Prize in Physiology or Medicine.

NOBEL PRIZE

Nobel announcement marred by winner's death

Immunology takes prize for medicine, but award comes three days too late for one recipient.

BY EWEN CALLAWAY

Before the immune system can attack an invading pathogen, it must identify the intruder. Breakthroughs in understanding this process have garnered three scientists this year's Nobel Prize in Physiology or Medicine, which was announced on 3 October in Stockholm.

But the award was quickly overshadowed by sadness when it emerged that the winner of half the 10-million-Swedish-krona (US\$1.5-million) prize, Ralph Steinman of the Rockefeller University in New York, had died on 30 September. Although Nobel prizes are not awarded posthumously, the Nobel committee was not aware of Steinman's death when it reached its decision. The committee has since confirmed that the award still stands. Ironically, Steinman was being treated for his pancreatic cancer with a therapy derived from his original discovery.

Together, his work and that of the other

two winners — Jules Hoffmann at the French National Centre for Scientific Research (CNRS) Institute of Cell and Molecular Biology in Strasbourg and Bruce Beutler of the Scripps Research Institute in La Jolla, California — help to describe how two separate arms of the immune system work.

Steinman discovered a type of immune cell, known as a dendritic cell, that is vital to the 'adaptive' immune system, which works out exactly which pathogen has invaded the body in order to trigger a targeted response. Hoffmann and Beutler earned their share of the prize for discovering a key to a more immediate line of defence, the 'innate' immune system, which identifies a foreign body as a potential pathogen. They identified the molecular sentinels that first sound the alarm by recognizing features shared by numerous pathogens.

Steinman's efforts to understand the immune system began in the early 1970s, when he joined the laboratory of Zanvil Cohn

at Rockefeller as a postdoc. Cohn's group was studying an immune cell called the macrophage, which engulfs pathogens and other debris. Most researchers thought that macrophages then alerted adaptive immune cells called T cells to the presence of a specific pathogen. Once activated, T cells multiply and combat infection, either by killing pathogen-infected cells or by steering another type of immune cell, the B cell, to produce pathogen-blocking antibodies.

In Cohn's lab, Steinman identified another type of immune cell, which he named the dendritic cell because of its long, tree-like arms¹. Cohn and Steinman showed that these cells are much more important than macrophages in activating T cells.

At first, dendritic cells "were a minor cell and everybody was loath to accept them", recalls Siamon Gordon, an immunologist at the University of Oxford, UK, who worked with Cohn and Steinman. "It was a bit like ►

► having two Pops — it was the dendritic cells versus the macrophages.” Steinman continued doggedly collecting data, and eventually won over his critics.

Two decades after the discovery of dendritic cells’ crucial role, a team led by Hoffmann was investigating why fruitflies, which lack an adaptive immune system, don’t succumb to fungal infection. In 1996, they reported that the *Toll* gene, previously linked to embryo development, was also important for battling infections². Flies with mutations in *Toll* died when exposed to bacteria or fungi.

At around the same time, a team led by Beutler, then at the University of Texas Southwestern Medical Center in Dallas, had spent six years looking for an immune-system gene in mice that produces a protein to recognize lipopolysaccharide (LPS), a molecule produced by certain bacteria that can cause septic shock. “We were obsessed,” says Alexander Poltorak, an immunologist now at Tufts University in Boston, Massachusetts, who worked on the project. “We always thought we would find the gene tomorrow.”

The team eventually found its LPS-sensing gene, and it looked remarkably like Hoffmann’s *Toll*³. Linking the two findings paved the way for the discovery of other Toll-like receptors that sense molecules made by pathogens but not their hosts, and form a critical part of the innate immune system.

The discoveries of dendritic cells and innate immune receptors have already had an impact on medicine. Vaccines are typically administered with an adjuvant, such as a metal, to prompt a rapid immune response. Drug companies such as Glaxo-SmithKline are now developing adjuvants that activate Toll-like receptors.

“By doing this we are mimicking what actually happens during an infection without having an infection,” says Vincenzo Cerundolo, associate director of the UK Medical Research Council Immunology Unit in Oxford.

Meanwhile, Provenge (made by the biotechnology company Dendreon of Seattle, Washington), the only cellular immune therapy against cancer to be approved by the US Food and Drug Administration, exploits dendritic cells that recognize a molecule produced by prostate tumours. Culturing and reinjecting the cells back into the patient fortifies the immune response against the tumour.

“The reason why the field has progressed so much and is now in the clinic is because we understand how to activate the immune system,” says Cerundolo. ■

1. Steinman, R. M. & Cohn, Z. A. *J. Exp. Med.* **137**, 1142–1162 (1973).
2. Lemaître, B. *et al. Cell* **86**, 973–983 (1996).
3. Poltorak, A. *et al. Science* **282**, 2085–2088 (1998).



Shining brightly: (from left) Saul Perlmutter, Brian Schmidt and Adam Riess.

NOBEL PRIZE

Stellar performance nets physics prize

Nobel for supernovae signals of accelerating Universe.

BY GEOFF BRUMFIEL

Three astrophysicists have been awarded a Nobel prize for planting a perplexing puzzle at the heart of cosmology.

Half of the Nobel Prize in Physics goes to Saul Perlmutter of Lawrence Berkeley National Laboratory in California for leading a team that discovered that the Universe is expanding at an ever-increasing rate (S. Perlmutter *et al. Astrophys. J.* **517**, 565–586; 1999). Brian Schmidt of the Australian National University in Weston Creek and Adam Riess of the Space Telescope Science Institute in Baltimore, Maryland, share the other half of the prize for independent measurements of the cosmic acceleration (A. G. Riess *et al. Astron. J.* **116**, 1009–1038; 1998), which researchers have struggled to explain ever since.

“I feel kind of weak in the knees,” Schmidt told reporters in Sweden via telephone. “It sort of feels like when my children were born.”

All three scientists reached their conclusions on the basis of measurements of distant Type Ia supernovae. These occur in very specific types of binary star system, in which a white dwarf star tears matter away from its partner until it gains enough mass to explode. At their peak, Type Ia supernovae always emit roughly the same amount of light, making them useful as ‘standard candles’ by which to measure vast distances across the cosmos.

In the late 1980s and early 1990s, the prize-winners precisely measured the brightness of these supernovae using newly developed digital sensors. They then compared the brightness to the redshift — the change in colour of the light that results from the motion of the supernovae away from us. Both teams found that the supernovae were dimmer than expected at the

measured redshift. The inescapable conclusion was that the Universe was not only expanding — which astronomers first realized in the 1920s — but expanding faster and faster.

Schmidt says that the finding was initially “pretty perplexing”. Most astronomers had expected that the Universe’s rapid growth following the Big Bang would gradually slow down as gravity pulled distant galaxies towards each other. Yet the discovery was accepted almost immediately by the astronomical community — in part because the idea of a cosmic pressure pushing the Universe outwards had already been mooted by Albert Einstein.

When Einstein applied his general theory of relativity to the Universe as a whole in 1917, his equations included a ‘cosmological constant’ which described just such an outward force. Over the past decade, observations of the large-scale structure of the Universe, together with the cosmic microwave background radiation — the faint afterglow of the Big Bang — have also indicated that the majority of the Universe’s energy remains undetected. Today, the astronomical community accepts that about 73% of the Universe’s energy is invested in this cosmic acceleration. Known as dark energy, it remains a largely mysterious force. “Nobody really knows what it is that has been discovered,” says Peter Coles, an astrophysicist at Cardiff University, UK.

The predominant view is that dark energy results from quantum fluctuations in the vacuum of space, but efforts to use quantum theory to describe it have so far failed. Other theories, including modifications of gravity, have gained little acceptance.

“It could be none of the above,” Coles says. But “we wouldn’t be on the trail of this ‘none-of-the-above’ if it hadn’t been for these experiments”. ■

► having two Pops — it was the dendritic cells versus the macrophages.” Steinman continued doggedly collecting data, and eventually won over his critics.

Two decades after the discovery of dendritic cells’ crucial role, a team led by Hoffmann was investigating why fruitflies, which lack an adaptive immune system, don’t succumb to fungal infection. In 1996, they reported that the *Toll* gene, previously linked to embryo development, was also important for battling infections². Flies with mutations in *Toll* died when exposed to bacteria or fungi.

At around the same time, a team led by Beutler, then at the University of Texas Southwestern Medical Center in Dallas, had spent six years looking for an immune-system gene in mice that produces a protein to recognize lipopolysaccharide (LPS), a molecule produced by certain bacteria that can cause septic shock. “We were obsessed,” says Alexander Poltorak, an immunologist now at Tufts University in Boston, Massachusetts, who worked on the project. “We always thought we would find the gene tomorrow.”

The team eventually found its LPS-sensing gene, and it looked remarkably like Hoffmann’s *Toll*³. Linking the two findings paved the way for the discovery of other Toll-like receptors that sense molecules made by pathogens but not their hosts, and form a critical part of the innate immune system.

The discoveries of dendritic cells and innate immune receptors have already had an impact on medicine. Vaccines are typically administered with an adjuvant, such as a metal, to prompt a rapid immune response. Drug companies such as Glaxo-SmithKline are now developing adjuvants that activate Toll-like receptors.

“By doing this we are mimicking what actually happens during an infection without having an infection,” says Vincenzo Cerundolo, associate director of the UK Medical Research Council Immunology Unit in Oxford.

Meanwhile, Provenge (made by the biotechnology company Dendreon of Seattle, Washington), the only cellular immune therapy against cancer to be approved by the US Food and Drug Administration, exploits dendritic cells that recognize a molecule produced by prostate tumours. Culturing and reinjecting the cells back into the patient fortifies the immune response against the tumour.

“The reason why the field has progressed so much and is now in the clinic is because we understand how to activate the immune system,” says Cerundolo. ■

1. Steinman, R. M. & Cohn, Z. A. *J. Exp. Med.* **137**, 1142–1162 (1973).
2. Lemaître, B. *et al. Cell* **86**, 973–983 (1996).
3. Poltorak, A. *et al. Science* **282**, 2085–2088 (1998).



Shining brightly: (from left) Saul Perlmutter, Brian Schmidt and Adam Riess.

NOBEL PRIZE

Stellar performance nets physics prize

Nobel for supernovae signals of accelerating Universe.

BY GEOFF BRUMFIEL

Three astrophysicists have been awarded a Nobel prize for planting a perplexing puzzle at the heart of cosmology.

Half of the Nobel Prize in Physics goes to Saul Perlmutter of Lawrence Berkeley National Laboratory in California for leading a team that discovered that the Universe is expanding at an ever-increasing rate (S. Perlmutter *et al. Astrophys. J.* **517**, 565–586; 1999). Brian Schmidt of the Australian National University in Weston Creek and Adam Riess of the Space Telescope Science Institute in Baltimore, Maryland, share the other half of the prize for independent measurements of the cosmic acceleration (A. G. Riess *et al. Astron. J.* **116**, 1009–1038; 1998), which researchers have struggled to explain ever since.

“I feel kind of weak in the knees,” Schmidt told reporters in Sweden via telephone. “It sort of feels like when my children were born.”

All three scientists reached their conclusions on the basis of measurements of distant Type Ia supernovae. These occur in very specific types of binary star system, in which a white dwarf star tears matter away from its partner until it gains enough mass to explode. At their peak, Type Ia supernovae always emit roughly the same amount of light, making them useful as ‘standard candles’ by which to measure vast distances across the cosmos.

In the late 1980s and early 1990s, the prize-winners precisely measured the brightness of these supernovae using newly developed digital sensors. They then compared the brightness to the redshift — the change in colour of the light that results from the motion of the supernovae away from us. Both teams found that the supernovae were dimmer than expected at the

measured redshift. The inescapable conclusion was that the Universe was not only expanding — which astronomers first realized in the 1920s — but expanding faster and faster.

Schmidt says that the finding was initially “pretty perplexing”. Most astronomers had expected that the Universe’s rapid growth following the Big Bang would gradually slow down as gravity pulled distant galaxies towards each other. Yet the discovery was accepted almost immediately by the astronomical community — in part because the idea of a cosmic pressure pushing the Universe outwards had already been mooted by Albert Einstein.

When Einstein applied his general theory of relativity to the Universe as a whole in 1917, his equations included a ‘cosmological constant’ which described just such an outward force. Over the past decade, observations of the large-scale structure of the Universe, together with the cosmic microwave background radiation — the faint afterglow of the Big Bang — have also indicated that the majority of the Universe’s energy remains undetected. Today, the astronomical community accepts that about 73% of the Universe’s energy is invested in this cosmic acceleration. Known as dark energy, it remains a largely mysterious force. “Nobody really knows what it is that has been discovered,” says Peter Coles, an astrophysicist at Cardiff University, UK.

The predominant view is that dark energy results from quantum fluctuations in the vacuum of space, but efforts to use quantum theory to describe it have so far failed. Other theories, including modifications of gravity, have gained little acceptance.

“It could be none of the above,” Coles says. But “we wouldn’t be on the trail of this ‘none-of-the-above’ if it hadn’t been for these experiments”. ■

MENTAL HEALTH

Trillion-dollar brain drain

Enormous costs of mental health problems in Europe not matched by research investment.

BY KERRI SMITH

Brain disorders cost Europe almost €800 billion (US\$1 trillion) a year — more than cancer, cardiovascular disease and diabetes put together. That's the conclusion of a report¹ commissioned by the European Brain Council that provides the most comprehensive assessment of the financial consequences of mental ailments so far.

The report's authors argue that these enormous costs — which exceed the entire gross domestic product of the Netherlands — mean that research into brain disorders receives disproportionately little funding compared with other diseases. They call on politicians and funders to step up support for basic research on these conditions, which are so costly because they often require long-term care and erode the productivity of those affected for years or decades.

The report is an update of a similar survey in 2005, which found that brain disorders were costing Europe €386 billion². Since then, Bulgaria and Romania have joined the European Union and seven more categories of disorder have been added to the assessment, including eating disorders, sleep disorders, mental retardation, and childhood and developmental disorders such as autism. The authors say that their new estimate, although double the 2005 figure, is likely to be “very conservative”.

Mood disorders top the cost estimates, consuming €113.4 billion a year, following closely by dementia, at €105.2 billion (see ‘Heavy burden’). On average, the annual cost per citizen is €1,550, with Luxembourg and the United Kingdom spending the most per head of population.

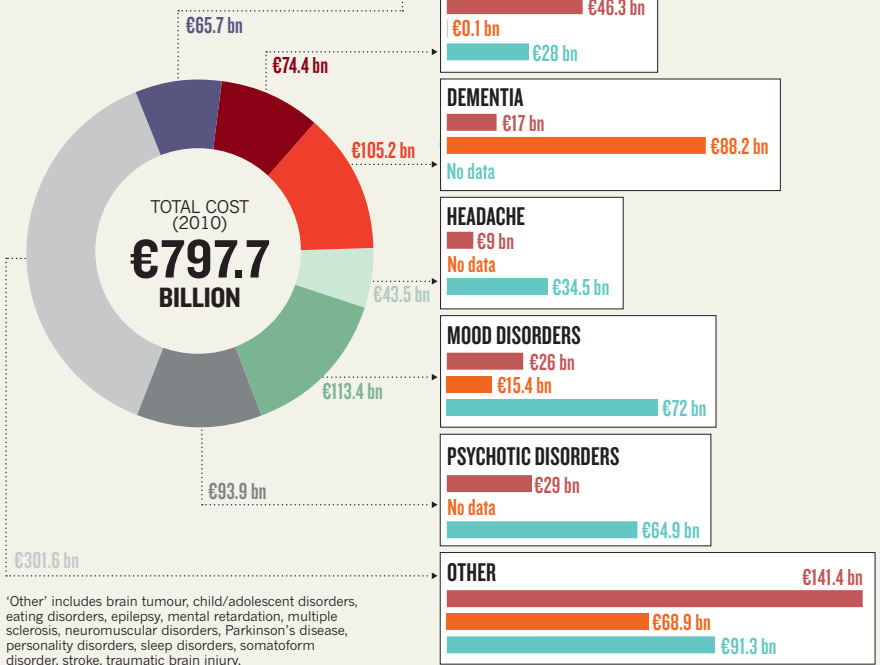
Drugs, visits to doctors and hospitalizations — the direct health-care costs — make up 37% of the bill. A further 23% is spent on direct non-medical costs, including informal care, social services and nursing homes. The remainder (40%) is sucked away by indirect costs, such as lost productivity as a result of time off work or early retirement. One reason for the high indirect costs is that “people don't tend to die quickly from brain disorders”, says Jes Olesen, the neurologist at the University of Copenhagen who headed the survey team. “People live for years in a disabled condition.”

➔ **NATURE.COM**
Grand challenges in
global mental health:
go.nature.com/wvj1The

More than 100 scientists and health economists in Europe were involved in collecting data for the report. For

HEAVY BURDEN

Six categories of illness account for more than half of the costs of brain disorders in Europe. Indirect costs — such as working time lost to illness — are responsible for about 40% of the total financial burden.



each country, the team found out how many people had a particular condition, estimated the financial costs, and then calculated Europe-wide figures. Where data weren't available, the prevalence and costs of disorders were estimated from figures from other countries. “They are of course imputations, but they are the best available,” says Olesen.

No directly comparable reports exist elsewhere in the world, but several studies have looked at the costs of individual conditions, such as bipolar disorder, attention deficit hyperactivity disorder and schizophrenia, in both Europe and the United States. Overall, health-care costs per person are similar in both regions, but the direct costs — doctors and drugs — are higher in America.

The drug industry, however, is increasingly shying away from these disorders. “The basic science is such that it's quite difficult to identify a new target, so you start with your hands tied behind your back,” says Patrick Vallance, head of medicines discovery and development for London-based drug giant GlaxoSmithKline, which last year stopped funding

drug-development programmes in psychiatry, pain and cognitive neuroscience. Vallance also cites problems with unrealistic animal models, unpredictable results from early trials and difficulties in diagnosing and allocating patients to trials. “At every stage of the process your risk is very much higher” for brain disorders than for other conditions, he says.

A report produced by the European Brain Council in 2006 (ref. 3) estimated that Europe spent about the same amount on brain research as on cancer research (about €4 billion each), despite the much higher cost of brain disorders.

Olesen says that the report presents clear evidence that greater scientific effort is required to tackle brain disorders. “The only way is to increase research and understand these disorders better,” says Olesen. Focusing on preventing these disorders in the first place would have the greatest cost benefit, he adds. ■

1. Gustavsson, A. *et al.* *Eur. Neuropsychopharmacol.* **21**, 718–779 (2011).
2. Andlin-Sobocki, P. *et al.* (eds) *Eur. J. Neurol.* **12**, Suppl. 1, 1–90 (2005).
3. Sobocki, P. *et al.* *Eur. J. Neurosci.* **24**, 2691–2693 (2006).

SOURCE: REF. 1

Europe cuts funding red tape

Changes to €80-billion Horizon 2020 research programme simplify grant process.

BY NATASHA GILBERT

It is not the revolution that some researchers hoped for. But *Nature* has learned that the next iteration of the European Union's multibillion-euro research programme includes modest yet significant reforms that could make the vast pot of funds easier for researchers to tap.

The European Commission acknowledges that burdensome administration, including a plethora of time sheets, financial audits and complicated rules, has discouraged some researchers from taking part in its €50.5-billion (US\$67.3-billion) Seventh Framework Programme, known as FP7. Plans for the next cycle of funding in 2014–20, called Horizon 2020, are not due for publication until late November, and although the commission has requested €80 billion for the programme, a budget has not yet been agreed. But *Nature's* early look at the plans reveals that Horizon 2020 will come with significantly less red tape.

The plans say that researchers will have to navigate only one set of rules for all initiatives in the programme, including collaborative research projects and those that fall under the auspices of the European Institute of Innovation and Technology, headquartered in Budapest. The rules will cover how proposals are evaluated, the criteria used to award funding and what indirect costs can be claimed for projects. Some variation in the rules will be allowed for issues such as the exploitation of research results and intellectual-property rights, in order to help innovation flourish — one of the key goals of Europe's commissioner for research, innovation and science, Máire Geoghegan-Quinn. "A common set of basic principles" across all the initiatives would lead to "considerable trimming and lightening of rules", the commission's proposals say.

"Any move to simplify the application and management of grants is to be welcomed," says Adam Hurlstone, a cancer researcher at the University of Manchester, UK. Hurlstone this year won a grant from the European Research Council (ERC), the pan-European initiative to fund frontier research solely on the basis of excellence, which has been lauded by researchers for its relatively simple application process. He says that he has previously been put off from applying for grants from other initiatives in the framework programme "due to my perception of the organizing and administrative burden".

The plans also say that the framework programme will be reorganized to group all



Ernst-Ludwig Winnacker (left) and European research commissioner Máire Geoghegan-Quinn have helped to reshape the Framework research funding programme.



P. GRIMM/PICTURE ALLIANCE/PHOTOSHOT; T. ROGE/REUTERS

'excellence' initiatives together under one administrative umbrella. This will include the ERC grants and the popular Marie Curie fellowships, which fund researchers to move from one EU country to another to work on a project. Unlike previous programmes, a portion of Horizon 2020 funding will be targeted towards six key challenges affecting society, including health, food security and clean energy.

Researchers should also find it easier to claim back the indirect costs of research projects, such as the provision of laboratory space and maintenance of equipment. The commission will sweep away the various ways of calculating indirect costs; instead, it proposes using a single rate for participants from institutions without commission-approved accounting systems. The move should help to avoid "frequent errors" in accounting, the commission says.

But the commission has decided not to eliminate the requirement for researchers to report in detail their research costs, meaning that they will still have to fill out time sheets and undergo financial auditing when projects are completed. It rejected an option in which researchers would be awarded all of their funding in a lump sum based on agreed research outputs. Instead, it settled for providing clearer advice to researchers filling in time sheets, and abolished time recording for researchers who work full time on a project, such as those with ERC grants.

The planning document says that more radical reforms would have required "major organisation changes in the commission", including building up new skills and reassigning staff to different roles. "I would like to have seen a more radical approach," says a senior European science adviser, who requested anonymity. "It is clear the commission is concerned for its own jobs rather than solving the problems that exist."

The Horizon 2020 proposals also rejected calls to give the ERC more independence and reduce the administrative burden imposed by the commission (see *Nature* <http://dx.doi.org/10.1038/news.2010.615>; 2010). Instead, the ERC will remain partially under the administrative control of the commission, as recommended in July by a task force including Helga Nowotny, the ERC president, and Ernst-Ludwig Winnacker, former secretary-general of the ERC. "In the past I would have recommended radical change, but we came to the conclusion that it would be too difficult and could threaten the ERC," says Winnacker.

The plans show that the commission is moving forward with other reforms suggested by the task force. These include replacing the commission's stifling daily supervision of the ERC with regular, but less frequent, meetings between the ERC's scientific leadership and a commission representative.

The commission will now refine the proposals and assign budgets before they are officially published. Discussions between the commission, member states and the European Parliament will take about 18 months. They hope to agree on a final version in time for the programme to begin in 2014. ■ **SEE EDITORIAL P.5**

ETHICS

Secrets of the human genome disclosed

Meeting debates ethics of revealing genetic findings.

BY ERIKA CHECK HAYDEN IN NEW YORK

Should people be told about any nasty surprises that scientists discover in their DNA during research projects?

The question is becoming increasingly pertinent, as thousands of people sign up for studies in which their genomes will be sequenced. But, at present, federal laws in the United States prohibit researchers from telling patients about mutations that might affect them or their families unless a certified clinical lab has confirmed the results — something that is not done in most research projects. This means that patients often do not learn about their mutations until the studies are finally published, a restriction that is meant to ensure they are not misinformed by incomplete research.

The ethical dilemmas became all too real last year for geneticist Gholson Lyon, a geneticist at the Utah Foundation for Biomedical Research in Salt Lake City. He was studying an extended family in which some of the boys had been born with a constellation of symptoms, including thick, wrinkly skin, and who ultimately died of cardiac disease before their first birthdays. By November 2010, Lyon had convincing evidence that a genetic mutation was causing the disease. That's when he learned that one of the women in the family was four months pregnant with a boy.

Lyon knew from his study that the mother carried the mutation. But he was not allowed to tell her, because the analysis had not been performed in a laboratory that was certified under the Clinical Laboratory Improvement Amendments, which aim to ensure that clinical tests are accurate and reliable.

The baby was eventually born with the disease — called Ogden syndrome — and later died, in the same week that Lyon's paper on the causative mutation was published¹.

At the fourth annual Personal Genomes meeting at Cold Spring Harbor in New York last week, Lyon argued that researchers should routinely conduct their studies in certified laboratories so that they can provide participants with results as soon as possible, adding that he plans to do so himself from now on. It is a pressing issue: according to Richard Gibbs, director of the Human Genome

Sequencing Center at Baylor College of Medicine in Houston, Texas, roughly 5,000

human genomes will be sequenced this year, with some 30,000 expected next year.

But ethicists point out that although researchers and physicians may feel obliged to disclose genetic information, they must also consider other factors. "This is not just about patients or doctors. These disclosures have societal implications that need to be considered, including downstream cost," says Ellen Wright Clayton, director of the Center for Biomedical Ethics and Society at Vanderbilt University in Nashville, Tennessee.

Genome sequencing is now starting to be used in the clinic to guide diagnosis and treatment decisions (see News Feature page 22).

"These disclosures have societal implications that need to be considered."

At the Medical College of Wisconsin in Milwaukee, for example, paediatrician and geneticist David Dimmock offers genome sequencing to children with undiagnosed diseases.

The programme is controversial because many researchers think that too little is known about how most rare genetic mutations contribute to disease for the knowledge to help patients. He points out, however, that a handful of cases have been reported in which sequencing has led to a cure or improved treatment².

Using a clinically certified lab, Dimmock's team sequenced the genome of an infant with acute liver failure, and discovered that she had two mutations in a gene called *Twinkle*. Earlier research had linked those mutations to progressive eye and neurological conditions, and an associated liver disease. As a result, doctors determined that a liver transplant — a standard treatment for acute liver failure — would not help the infant, and recommended against it³. She died when she was 6 months old.

"This was not a happy ending — but in a sense it was," says Dimmock. Disclosing the genetic information spared the infant from spending her remaining few months recovering from a gruelling, unnecessary transplant, he says, and saved a scarce liver for a child who might benefit from it more. ■

1. Rope, A. F. *et al.* *Am. J. Hum. Genet.* **89**, 28–43 (2011).
2. Bainbridge, M. N. *et al.* *Sci. Transl. Med.* **3**, 87re3 (2011).
3. Goh, V. *et al.* *J. Pediatr. Gastroenterol. Nutr.* <http://dx.doi.org/10.1097/mpg.0b013e318227e53c> (2011).

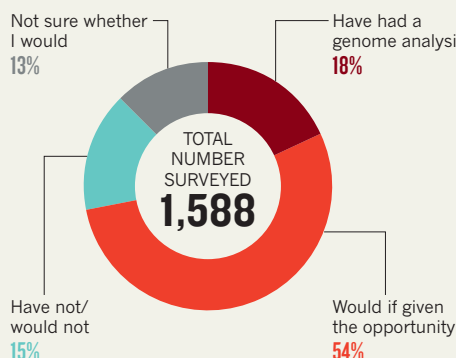
► NATURE.COM
The human genome at 10
nature.com/humangenome

GENOMES BY THE NUMBERS

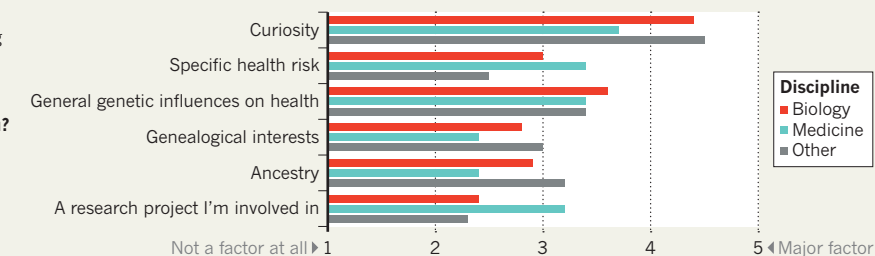
Of 1,588 respondents, 289 report having taken a total of 396 genetic tests, ranging from whole-genome sequencing to testing of a single gene.

For full poll results, visit go.nature.com/9ihtf2

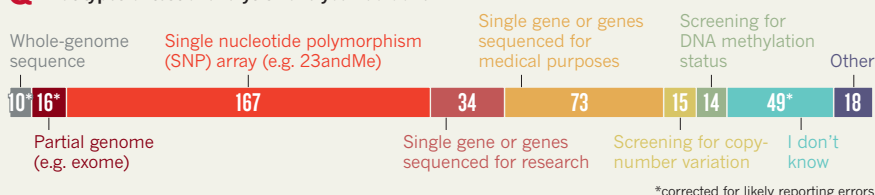
Q Have you had a genome analysis and if not, would you?



Q How important were the following factors in influencing your decision to have a genome analysis done?



Q What types of test or analysis have you had done?



GENE SEQUENCING

Nature readers flirt with personal genomics

Survey reveals eagerness to use latest DNA technologies.

BY BRENDAN MAHER

Ronald Worthington lives an examined life. The genome analyst, who works at Southern Illinois University School of Pharmacy in Edwardsville, got funding from his university to have his entire genome sequenced by Complete Genomics in Mountain View, California. He also ordered the sequence of the coding regions of his genome, known as the exome, from Otogenetics in Tucker, Georgia, and he will soon be giving skin and blood samples that will be used to immortalize his cells and DNA. He aims to contribute it all to the Personal Genome Project, an ambitious effort to sequence 100,000 individuals and post their data and medical histories online for anyone to access.

"This approach is the way to jump-start this whole process of integrating human genomic data into clinical medicine," he says.

Worthington is one of 1,588 people who responded last month to a *Nature* poll on readers' attitudes to personal genomics (see 'Genomes by the numbers'). Participants were recruited by e-mail, and through *Nature's* online Facebook and Twitter accounts, the Nature News Blog and Genomes Unzipped, an independent blog that chronicles developments in personal genomics. It seems that, overall, *Nature* readers are eager to adopt these new technologies. About 18% report

having had their genomes analysed in some way, ranging from whole-genome sequencing (about 10 respondents, after correcting for reporting errors) to direct-to-consumer tests. Of the remainder, 66% say they would have their genome sequenced or analysed if the opportunity arose.

Although scientists dominated our sample, only some 20% of those whose genomes had been analysed reported that their research goals or those of others were a major factor in the decision. About 30% did some or all of the analysis themselves. Worthington, for example, isolated his DNA for sequencing and is annotating his exome sequence. But 50% used the services of 23andMe of Mountain View, California, a DNA-testing company that surveys each customer's genome for 1 million of the DNA markers known as SNPs to trace ancestry and to predict disease risks.

"When I found out about direct-to-consumer genetic testing, I thought, 'I've got to do this'."

Worthington, too, bought a 23andMe kit and says that the results allayed his anxieties about having his full genome sequenced after they revealed no susceptibility to serious illnesses. He also notes the naivety of his rationale, saying: "It is easily possible that next week a genome-wide association study will report that, based upon my 23andMe genotypes alone,

I am at substantial risk for some as yet under-appreciated terrible disease."

But for other respondents, particularly those in medicine or public health, probing disease risk was a primary motivation. Kelly Leight, who coordinates the group Preserving the Future of Newborn Screening, based in Short Hills, New Jersey, became an advocate of personal genomics after her daughter was diagnosed with late-onset congenital adrenal hyperplasia, a genetic disorder that can be serious if untreated. Leight, her husband and her daughter were all sequenced for the causal gene to confirm the diagnosis. Later, she learned about 23andMe. "When I found out about direct-to-consumer genetic testing, I thought, 'This is totally for me. I've got to do this.'"

Like many poll respondents, she says her genome is now more of a hobby than anything else. Health-risk information from a genome scan did persuade her to lose 60 pounds in weight, putting her among the 27% of those in the poll who changed their behaviour because of information in their genome. She's spread the word to her family, buying genotyping kits for family members. She says that her geneticist and genetic-counsellor friends are shocked that she would push such a personal decision on her family. But Leight sees no harm. Some family members embraced their genomes, others ignored them. But "I didn't put any pressure on them", Leight says. "Well, maybe I did." ■

METRICS

Experts question rankings of journals

F1000 scoring system could throw off results, say critics.

BY DECLAN BUTLER

Peer review may be a good way to assess research papers, but it can fall short in ranking the journals themselves. That's the reaction of some metrics experts to the first such journal rankings, launched this week by the Faculty of 1000 (F1000) in London. Critics question the method, which relies on scores awarded to individual papers by the F1000 'faculty' of 10,000 scientists and clinicians. Such scores, they claim, could be skewed by the interests and enthusiasms of individual reviewers.

Richard Grant, associate editor of F1000, says that the rankings give authors a valuable measure, complementary to journal league tables based on citation impact. He says that the first F1000 rankings will be refined, adding that F1000 is "constantly striving to

improve coverage in all specialities".

Created in 2002, the F1000 aims to filter the literature by asking experts to select noteworthy papers and rate them as 6 (recommended), 8 (must read) or 10 (exceptional). Now, it has extended the concept by totting up the scores of all a journal's rated articles over a given period, and normalizing the totals — adjusting for the total number of articles that the journal published over that period, for example.

The results put the usual suspects at the top. In the rankings for 2010, the latest full year available, *Nature* leads in biology and the *New England Journal of Medicine* in medicine. But further down the lists, the F1000 often departs from impact factors. "We're aware the correlation with impact factors isn't exact, and we wouldn't expect it to be," says Grant. The *Proceedings of the National Academy of Sciences*

(*PNAS*) "does particularly well by our ranking because there are a lot of papers in there that are obviously valuable to the community".

But some critics say that the limited number of papers reviewed — fewer than 20,000 per year, of more than one million published — could compromise the rankings. "The scores may tell us as much about the composition of the F1000 faculty as they do about the relative quality of various journals," says Carl Bergstrom, a biologist at the University of Washington, Seattle, and an F1000 faculty member who publishes a rival metric, the Eigenfactor.

Philip Davis, a scholarly-publishing expert at Cornell University in Ithaca, New York, says that "a single enthusiastic reviewer could propel a small, specialist journal into a high ranking simply by submitting more reviews". One journal seems to owe its surprisingly high ranking to a series of very positive evaluations of its articles by its own editor.

Grant says that such a competing interest should have been declared, and that the F1000 will look into the matter. In the interim, the journal in question has been withdrawn from the rankings. The F1000 will also make its code of conduct more explicit, says Grant. He notes that all evaluations and methodology are available on the F1000's website, making the ranking process transparent and allowing users to alert the F1000 to any concerns. ■

GENOMES ON PRESCRIPTION

The first clinical uses of whole-genome sequencing show just how challenging it can be.

BY BRENDAN MAHER

The first thing Debbie Jorde noticed about her newborn daughter was that her arms were bent at unnatural angles. She had other problems, too: a cleft palate, eight fingers, eight toes and no lower eyelids. She would eventually be diagnosed with Miller syndrome, a disease so rare that doctors have long assumed that each case arises through spontaneous mutation, rather than being passed down through families. Doctors told Jorde that her chances of having a second child with the syndrome were less than one in a million.

They were wrong. Jorde's son, born three years after his sister, had the same features. Lynn Jorde, Debbie's current husband and a geneticist at the University of Utah in Salt Lake City, still cringes when Debbie recounts what the doctors had told her. "The right answer for that situation is that there have been so few cases that we really can't predict the risk," he says.

Thanks to next-generation genome sequencing, Debbie and her children now know the family's genetic risks. Lynn and his collaborators had been talking about sequencing the genomes of an entire 'nuclear' family affected by a genetic disease, both to identify the mutation responsible and to investigate how genes are inherited in unprecedented detail. Debbie, her former husband and her now-adult children, Heather and Logan Madsen, were happy to be take part, and in 2009 became the first family in the world to have their genomes fully sequenced¹.

Over the course of six months, the research team cross-compared the whopping amount of DNA data from the four genomes. With the help of a parallel sequencing effort that included others



with Miller syndrome², the researchers identified the gene involved, called *DHODH*, which encodes a protein involved in the synthesis of nucleotides. The disease, it turns out, is recessive. In this case, both parents carried a single mutated copy of the gene, so their chance of having a child with the syndrome was actually one in four. The analyses also revealed that the children had a second recessive genetic disorder, primary ciliary dyskinesia, which affects lung development. Before that discovery, says Debbie, “We never knew why they kept getting pneumonia.”

Families like Debbie Jorde’s are part of a small but growing vanguard of people, mostly with rare diseases and cancers, whose genomes have been sequenced to help diagnose or understand their condition. Although knowing the sequence didn’t alter treatments for Heather and Logan, some individuals are being sequenced with that intent. A boy in Wisconsin was given a risky but life-saving bone-marrow transplant last year on the basis of a partial genome sequence³; a woman with leukaemia was spared a similar procedure after her genome was sequenced⁴; and genome sequencing was used to refine the therapy given to twins with a rare disorder (see ‘6 billion to one’)⁵.

Most of those involved so far have been lucky enough to know the right people — researchers with an interest in clinical genetics — or determined enough to seek them out, and many, such as Debbie Jorde’s family, were taking part in research projects. But now, with genome sequencing becoming much cheaper and faster, clinical programmes are starting up around the world that will routinely analyse genomes for those who might benefit from the information. Illumina, which is based in San Diego, California, and provided the sequencing machines for many of the programmes, offers whole-genome sequencing for as little as US\$7,500 for people with life-threatening disease, and for \$10,000 for people with cancers that require the sequencing of both tumour and non-cancer cells.

As prices fall further, some say that prescribing a genome sequence or analysis will become akin to requesting a magnetic resonance imaging (MRI) scan. “It’s just like any other test in medicine. There’s nothing remotely special about it,” says David Bick, a clinical geneticist at the Medical College of Wisconsin in Milwaukee. But, he adds, “people will cry and scream and yell about that statement”. That’s true: unlike the results of most medical tests, a genome sequence provides a vast amount of difficult-to-interpret data, not all of which will be necessary for diagnosing or treating the patient’s condition and which could provide unwanted clues to future health risks. The few success stories published so far also suggest that wringing information from the human genome and counselling patients and their families adequately may be too big a burden for medical systems that are already stretched to their limits. “You can’t immediately jump from those few profound but limited stories and think that you can reduce this to practice for clinical care,” says Eric Green, director of the National Human Genome Research Institute (NHGRI) in Bethesda, Maryland. Still, from the pioneering cases, much can be learned.

RARE BIRTHS

Take Nicholas Volker. From the time he was born, an undiagnosed condition ravaged his intestines, sometimes causing fistulae: holes that ran from his gut through to the outside of his body, leaking faeces and requiring surgery. By the time he turned three, Volker had been in an operating room more than 100 times. Doctors hypothesized that he had an immune deficiency and that a bone-marrow transplant might correct the problem. But a number of tests, including the sequencing of several genes, were inconclusive. After intense deliberation, a team at the Medical College of Wisconsin was cleared to sequence Volker’s exome, the 1–2% of the genome that codes for proteins and key regulatory RNA molecules.

Using computational tools, the team combed Volker’s DNA for sequences that vary from person to person. They compared these with known variants in the general population, with variants associated with diseases and with related sequences in other species, looking for a mutation that might have caused the problem, says David Dimmock, a clinical geneticist at the college. It took, “basically one person staring at a computer for three and a half months”, he says, but eventually they identified

a mutation on the X chromosome in a gene called *X-linked inhibitor of Apoptosis*, or *XIAP* (ref. 3). A deficiency of the protein encoded by this gene is known to put patients at high risk for a deadly immune-cell disorder, and a bone-marrow transplant suddenly became imperative. More than a year later, Dimmock says, Volker is doing well.

What started as an experiment has become a programme at Wisconsin, where Dimmock, Bick and their colleagues now aim to provide comprehensive whole-genome sequencing for patients. The team is focusing on people with rare disorders that are thought to involve a genetic defect, and in whom identifying that defect is likely to inform the course of treatment.

Bick says that of 48 patients evaluated for the programme, 17 have been accepted, and their families have gone through six hours or more of genetic counselling before sequencing. Insurance companies have agreed to foot the bill for at least two of the cases. Their rationale is straightforward, says Tina Hambuch, a senior scientist at Illumina’s clinical services laboratory, which has been doing the sequencing for this programme. A full genome sequence can be less expensive than a series of single genetic tests, and might clarify whether a costly treatment is required. “There are cases where it’s cost effective,” Hambuch says.

GENOME FACTORIES

Other institutions are following suit. In the United Kingdom, the Wellcome Trust Centre for Human Genetics at the University of Oxford has made plans with Illumina to sequence 500 genomes from people — some from the same family — with a wide range of conditions. The Undiagnosed Diseases Program at the National Institutes of Health in Bethesda has been running a sequencing programme since 2008. It has sequenced more than 140 exomes and 5 genomes in its attempts to find the molecular underpinnings of diseases that have eluded diagnosis. The programme was so overwhelmed by interest that it temporarily stopped accepting applications a few months ago.

Green says that “now is the time to push the accelerator”. Clinical geneticists often talk about tackling Mendelian disorders: diseases thought to involve a single gene and that roughly obey the rules of inher-

“WE’VE LEARNED A LOT ABOUT HOW HARD EVALUATING AN EXOME IS.”

itance drawn up by Gregor Mendel in the nineteenth century. These conditions may account for as many as 20% of paediatric hospitalizations worldwide and a large share of health-care costs. Yet their genetic basis is often unknown. The compendium of such conditions, called Online Mendelian Inheritance in Man (OMIM), currently contains just under 7,000 disorders, about half of which have been assigned a molecular cause. This autumn, Green says, the NHGRI will announce the winners of its Mendelian Disorders Genome Centers grants, which will fund sequencing centres looking for causes of the rest.

Still, many researchers worry that it will be difficult to make clinical use of most genomes. At the Undiagnosed Diseases Program, the misses have certainly outnumbered the hits so far. “I think we’ve learned a lot about how hard evaluating an exome is,” says Thomas Markello, from the medical-genetics branch of the NHGRI. “I’m most concerned that people don’t recognize that what’s been published to date are the success stories.”

Many researchers say that genome sequencing could be used in diagnosis and therapy of cancer more easily than in rare diseases. Clinicians are already doing sophisticated analyses of some tumours in order to tailor therapies to the patient’s genetic characteristics; a genome sequence provides even more molecular detail. For example, an individual’s cancer genome sometimes reveals defects in a pathway that might point to use of a known drug, but were not apparent from standard tests.

In 2007, a 78-year-old man in Canada with a rare tongue cancer that had spread throughout his body was being treated at the British Columbia

6 BILLION TO ONE

Sequencing a genome is simple; finding the genetic cause of a disease is not. In one study⁵, scientists sequenced the genomes of fraternal twins diagnosed with a movement disorder and whittled down more than 1.6 million variants shared by the twins to the mutations in just one gene. The molecular diagnosis led to improved treatment for the twins.

6 billion nucleotides
Diploid human genome

1.63 million Single-base variants shared by twins that differ from reference human genome

9,531 Variants that code for proteins

4,605 Variants that change amino-acid sequence

77 Rare variants (which are more likely to cause disease)

3 Candidate genes

One gene linked to disorder

Cancer Agency in Vancouver. There was no approved treatment for his type of cancer and — being what doctors described as a “savvy sort” — he and his clinician convinced scientists at the agency to sequence the cancer’s genome. The scientists also analysed its transcriptome, revealing both the sequence and the amount of RNA that the tumour was producing. The team then compared these data with those for other cancers and for the patient’s normal cells.

The researchers homed in on *RET*, a gene known to promote cancer, which was duplicated in the tumour genome and churning out RNA. Several drugs are known to inhibit the protein encoded by this gene. Marco Marra, director of the cancer agency’s genome-sciences centre, says that “after much agonizing and hand-wringing”, the clinical team prioritized these drugs and tried the top one, sunitinib. The cancer stabilized for several months on this and a second treatment, but eventually started to spread again. An analysis of the recurring tumours showed that different cancer-promoting pathways had been activated⁶, making the tumours resistant to the first drug, but possibly responsive to others. Unfortunately, by then it was too late to do more: the man died.

UNSTOPPABLE TRAIN

Marra’s group is now setting up a project to better diagnose subtypes of another cancer, acute myelogenous leukaemia, using transcriptome and other sequencing methods. Partly inspired by Marra’s efforts, Elaine Mardis, a geneticist at Washington University in St Louis, Missouri, and her collaborators have used genome sequencing to try to help a handful of people with cancer, including the woman with leukaemia⁴. The woman had been treated and had gone into remission, but standard tests were unable to show conclusively whether she had acute promyelocytic leukaemia (APL) — which generally has a good outcome with standard therapy — or a type of leukaemia that would require aggressive follow-up treatment, such as a bone-marrow transplantation. Over about seven weeks, the team sequenced the cancer’s genome and found a gene fusion that was consistent with APL. Mardis is enthusiastic about the approach, but notes its limitations. “It’s another piece of evidence,” she says. “It’s not going to be the only thing that you’re looking at when going to a patient diagnosis.”

Moving whole-genome sequencing from research to clinic is beset with challenges. Unlike in research, DNA sequencing that is intended to inform a diagnosis must be done in accredited laboratories, such as those

used by Illumina. The institutional review boards that oversee research in humans have not reached a consensus on whether approval is needed for clinical genome sequencing; and the US Food and Drug Administration is yet to work out how to regulate the coming wave of clinical sequencing.

Many researchers and clinicians worry that health systems don’t have enough people well versed in genomics or bioinformatics to interpret the flood of data. What’s more, say experts, function and disease information for the human genome is scattered across scientific articles and databases that are hard to troll through and aren’t always correct. Sequence analysis is where most costs now lie. Hambuch says that for the few research projects on which Illumina has collaborated, just identifying all the variants in a genome has taken two to three weeks. “That’s a lot of effort from high-skill people,” she says.

The information could also overwhelm patients. Medical geneticists and ethicists have long worried about finding genetic pointers to disease risks that are unrelated to the illness being treated. With a full genome sequence, the likelihood of such incidental findings shoots up. The situation is particularly tricky for young patients. Do parents have the right to decide for them what information is revealed? This is where many of those hours of genetic counselling are spent, says Bick.

For these reasons, Stephen Kingsmore at Children’s Mercy Hospital in Kansas City, Missouri, argues that clinical sequencing should be limited in scope. He advocates sequencing just what he calls the Mendelianome, the genetic regions known to be involved in inherited diseases. “Ethically, legally, socially that’s going to be more acceptable,” he says. His group is developing methods that use a panel of mutations associated with just over 600 recessive diseases for such screening. Doing much more than this, he says, puts research goals ahead of the patient.

But some geneticists think that the train is unstoppable. “Once you demonstrate how informative this technology is, I think this is going to be a widely adopted,” says Hakon Hakonarson, who is starting a programme for clinical assessment of genomes at the Children’s Hospital of Philadelphia in Pennsylvania.

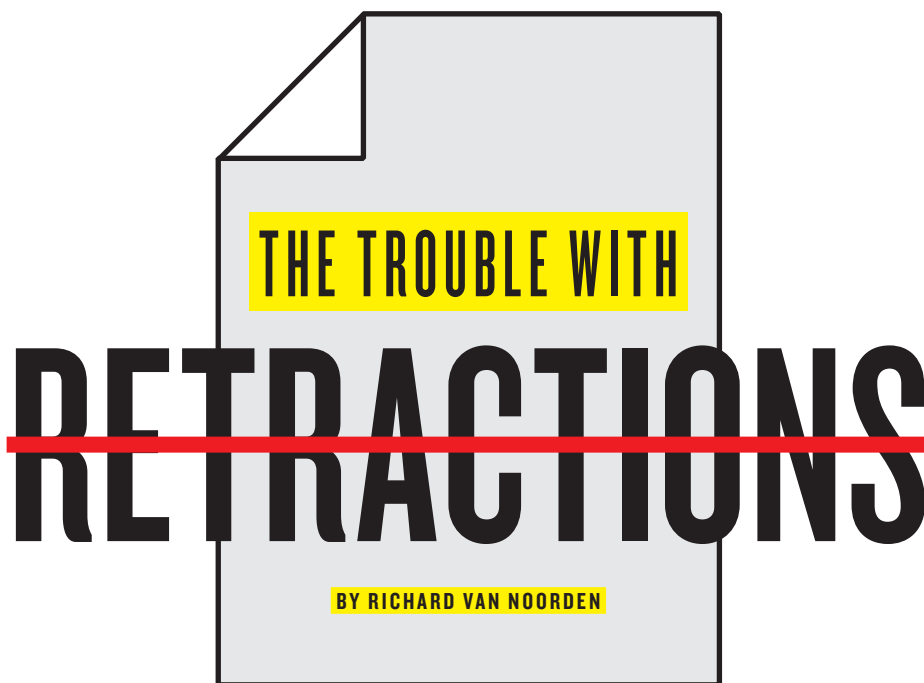
The members of Debbie Jorde’s family still ponder what their genome sequences have meant for them. Although the sequences didn’t alter treatment, if they had known about the lung problem earlier it might have prevented a dangerous procedure that both Heather and Logan underwent to reduce the recurrence of pneumonia.

Still, Lynn Jorde thinks that more successes are on the way for genomes in clinical care. “I’d predict some spectacular applications.” But, he adds, “I’m a congenital optimist”. ■ **SEE NEWS P.17 & P.19**

Brendan Maher is a features editor for Nature.

1. Roach, J. C. *et al.* *Science* **328**, 636–639 (2010).
2. Ng, S. B. *et al.* *Nature Genet.* **42**, 30–35 (2010).
3. Worthey, E. A. *et al.* *Genet. Med.* **13**, 255–262 (2010).
4. Welch, J. S. *et al.* *J. Am. Med. Assoc.* **305**, 1577–1584 (2011).
5. Bainbridge, M. N. *et al.* *Sci. Transl. Med.* **3**, 87re3 (2011).
6. Jones, S. J. M. *et al.* *Genome Biol.* **11**, R82 (2010).

➔ **NATURE.COM**
To hear more
about genomes on
prescription:
go.nature.com/pdcrsh



THE TROUBLE WITH RETRACTIONS

BY RICHARD VAN NOORDEN

A surge in withdrawn papers is highlighting weaknesses in the system for handling them.

This week, some 27,000 freshly published research articles will pour into the Web of Science, Thomson Reuters' vast online database of scientific publications. Almost all of these papers will stay there forever, a fixed contribution to the research literature. But 200 or so will eventually be flagged with a note of alteration such as a correction. And a handful — maybe five or six — will one day receive science's ultimate post-publication punishment: retraction, the official declaration that a paper is so flawed that it must be withdrawn from the literature.

It is reassuring that retractions are so rare, for behind at least half of them lies some shocking tale of scientific misconduct — plagiarism, altered images or faked data — and the other half are admissions of embarrassing mistakes. But retraction notices are increasing rapidly. In the early 2000s, only about 30 retraction notices appeared annually. This year, the Web of Science is on track to index more than 400 (see 'Rise of the retractions') — even though the total number of papers published has risen by only 44% over the past decade.

Perhaps surprisingly, scientists and editors broadly welcome the trend. "I don't think there's any doubt that we're detecting more fraud, and that systems are more responsive to misconduct. It's become more acceptable for journals to step in," says Nicholas Steneck, a research ethicist at the University of Michigan in Ann Arbor. But as retractions become more commonplace,

stresses that have always existed in the system are starting to show more vividly.

When the UK-based Committee on Publication Ethics (COPE) surveyed editors' attitudes to retraction two years ago, it found huge inconsistencies in policies and practices between journals, says Elizabeth Wager, a medical writer in Princes Risborough, UK, who is chair of COPE. That survey led to retraction guidelines that COPE published in 2009. But it's still the case, says Wager, that "editors often have to be pushed to retract".

Other frustrations include opaque retraction notices that don't explain why a paper has been withdrawn, a tendency for authors to keep citing retracted papers long after they've been red-flagged (see 'Withdrawn papers live on') and the fact that many scientists hear 'retraction' and immediately think 'misconduct' — a stigma that may keep researchers from coming forward to admit honest errors.

Perfection may be too much to expect from any system that has to deal with human error in all its messiness. As one journal editor told Wager, each retraction is "painfully unique".

But as more retractions hit the headlines, some researchers are calling for ways to improve their handling. Suggested reforms include better systems for linking papers to their retraction notices or revisions, more responsibility on the part of journal editors and, most of all, greater transparency and clarity about mistakes in research.

The reasons behind the rise in retractions are still unclear. “I don’t think that there is suddenly a boom in the production of fraudulent or erroneous work,” says John Ioannidis, a professor of health policy at Stanford University School of Medicine in California, who has spent much of his career tracking how medical science produces flawed results.

In surveys, around 1–2% of scientists admit to having fabricated, falsified or modified data or results at least once (D. Fanelli *PLoS ONE* 4, e5738; 2009). But over the past decade, retraction notices for published papers have increased from 0.001% of the total to only about 0.02%. And, Ioannidis says, that subset of papers is “the tip of the iceberg” — too small and fragmentary for any useful conclusions to be drawn about the overall rates of sloppiness or misconduct.

Instead, it is more probable that the growth in retractions has come from an increased awareness of research misconduct, says Steneck. That’s thanks in part to the setting up of regulatory bodies such as the US Office of Research Integrity in the Department of Health and Human Services. These ensure greater accountability for the research institutions, which, along with researchers, are responsible for detecting mistakes.

The growth also owes a lot to the emergence of software for easily detecting plagiarism and image manipulation, combined with the greater number of readers that the Internet brings to research papers. In the future, wider use of such software could cause the rate of retraction notices to dip as fast as it spiked, simply because more of the problematic papers will be screened out before they reach publication. On the other hand, editors’ newfound comfort with talking about retraction may lead to notices coming at an even greater rate.

“Norms are changing all the time,” says Steven Shafer, editor-in-chief of the journal *Anesthesia & Analgesia*, who has participated in two major misconduct investigations — one of which involved 11 journals and led to the retraction of some 90 papers.

IT’S NONE OF YOUR DAMN BUSINESS!

But willingness to talk about retractions is hardly universal. “There are a lot of publishers and a lot of journal editors who really don’t want people to know about what’s going on at their publications,” says New York City-based writer Ivan Oransky, executive editor at Reuters Health. In August 2010, Oransky co-founded the blog Retraction Watch with Adam Marcus, managing editor at *Anesthesiology News*. Since its launch, Oransky says, the site has logged 1.1 million page views and has covered more than 200 retractions.

In one memorable post, the reporters describe ringing up one editor, L. Henry Edmunds at the *Annals of Thoracic Surgery*, to ask about a paper withdrawn from his

journal (go.nature.com/ubv261). “It’s none of your damn business!” he told them. Edmunds did not respond to *Nature*’s request to talk for this article.

The posts on Retraction Watch show how wildly inconsistent retractions practices are from one journal to the next. Notices range from informative and transparent to deeply obscure. A typically unhelpful example of the genre would be: “This article has been withdrawn at the request of the authors in order to eliminate incorrect information.” Oransky argues that such obscurity leads readers to assume misconduct, as scientists making an honest retraction would, presumably, try to explain what was at fault.

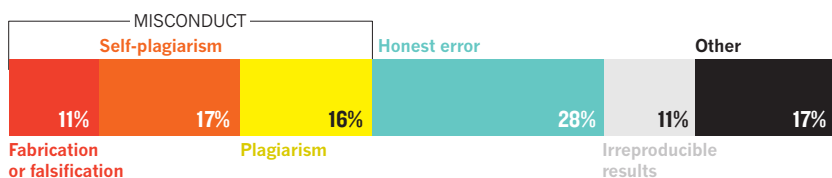
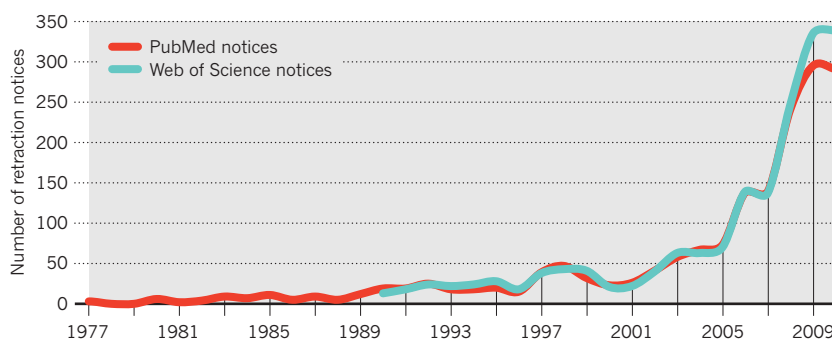
To Drummond Rennie, deputy editor of the *Journal of the American Medical Association*, there are two obvious reasons for obscure retraction notices: “fear and work.”

The fear factor, says Wager, is because publishers are very frightened of being sued. “They are incredibly twitchy about publishing anything that could be defamatory,” she says.

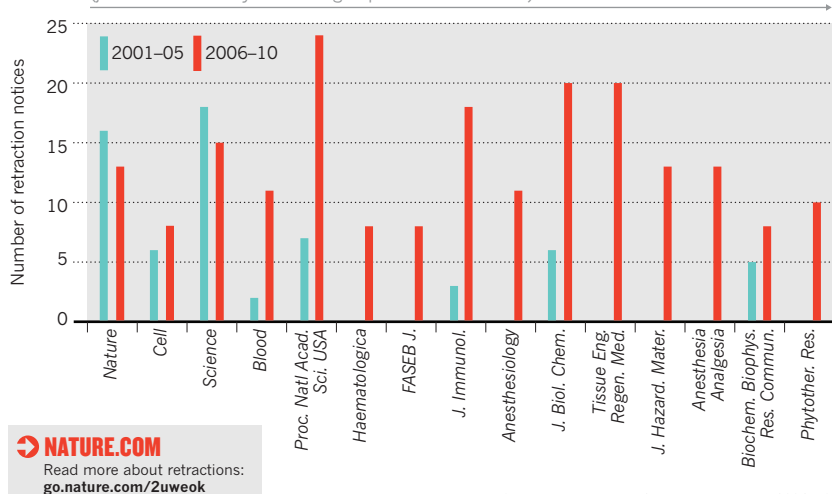
‘Work’ refers to the phenomenal effort required to sort through authorship disputes, concerns about human or animal subjects, accusations of data fabrication and all the other ways a paper can go wrong. “It takes dozens or hundreds of hours of work to get to the bottom of what’s going on and really understand it,” says Shafer. Because most journal editors

RISE OF THE RETRACTIONS

In the past decade, the number of retraction notices has shot up 10-fold (top), even as the literature has expanded by only 44%. It is likely that only about half of all retractions are for researcher misconduct (middle). Higher-impact journals have logged more retraction notices over the past decade, but much of the increase during 2006–10 came from lower-impact journals (bottom).



JOURNALS WITH MORE THAN 7 RETRACTION NOTICES IN WEB OF SCIENCE*, 2006–10 (journals ordered by decreasing impact factor for 2010)



NATURE.COM

Read more about retractions:
go.nature.com/Zuweek

*Not shown: *Acta Crystallographica E* saw 81 retractions during 2006–10.

are scientists or physicians working on a voluntary basis, he says, that effort comes out of their research and clinical time.

But the effort has to be made, says Steneck. “If you don’t have enough time to do a reasonable job of ensuring the integrity of your journal, do you deserve to be in business as a journal publisher?” he asks. Oransky and Marcus have taken a similar stance. This summer, for example, Retraction Watch criticized the *Journal of Neuroscience* for a pair of identical retraction notices it published on 8 June: “At the request of the authors, the following manuscript has been retracted.”

But the journal’s editor-in-chief, neuroscientist John Maunsell of Harvard Medical School in Boston, Massachusetts, argues that such obscurity is often the most responsible course to take. “My feeling is that there are far fewer retractions than there should be,” says Maunsell, who adds that he has conducted 79 ethics investigations in more than 3 years at the journal — 1 every 2–3 weeks. But “authors are reluctant to retract papers,” he says, “and anything we put up in the way of a barrier or disincentive is a bad thing. If authors are happier posting retractions without extra information, I’d rather see that retraction go through than provide any discouragement.”

At the heart of these arguments, says Steneck, lie shifting norms of how responsible journal editors should be for the integrity of the research process. In the past, he says, “they felt that institutions and scientists ought to do it”. More and more journal editors today are starting to embrace the gatekeeper role. But even now, Shafer points out, they have only limited authority to challenge institutions that are refusing to cooperate. “I have had institutions, where I felt there was very clear misconduct, come back and tell me there was none,” Shafer says. “And I have had a US institution tell me that they would look into allegations of misconduct only if I agreed to keep the results confidential.”

THE BLAME GAME

Discussions on Retraction Watch make it clear that many scientists would like to separate two aspects of retraction that seem to have become tangled together: cleaning up the literature, and signalling misconduct. After all, many retractions are straightforward and honourable. In July, for example, Derek Stein, a physicist at Brown University in Providence, Rhode Island, retracted a paper in *Physical Review Letters* on DNA in nanofluidic channels when he found that a key part of the analysis had been performed incorrectly. His thoroughness and speed — the retraction came just four months after publication — were singled out for praise on Retraction Watch.

But because almost all of the retractions that hit the headlines are dramatic examples of misconduct, many researchers assume that any retraction indicates that something shady has

DO RETRACTIONS WORK?

Withdrawn papers live on

In theory, retracting a paper is tantamount to withdrawing it from the scientific literature, so that it will never again mislead anyone. But when John Budd, at the School of Education at the University of Missouri in Columbia, examined 235 articles retracted during 1966–96, he found that they were cited in total more than 2,000 times after their withdrawal, with fewer than 8% of the citations acknowledging the retraction. And the rates haven’t improved much in the age of electronic publication: in a preliminary analysis of 1,112 retracted papers during 1997–2009, Budd finds them cited just as often, with the retraction mentioned in only about 4% of the citations. Other studies suggest that the situation is even worse for corrections, which are more numerous and often add important updates to a paper.

One solution is being developed by CrossRef, a non-profit collaboration of 3,599 commercial and learned-society publishers. It tries to address the fact that many researchers today never see corrections or retraction notices because they just download digital, PDF-formatted copies of the papers they need, and never again consult the original source. A new system, called CrossMark, consists of a logo that publishers will put on every PDF. Clicking on the logo will show Internet-connected users any updates to the work, whether retractions, corrections or other notes. The project is expected to launch in early 2012.

That will help researchers become aware of updates that have been recorded. But most science doesn’t progress by revising its written record. Papers superseded by later work, or that are controversial for some reason, are usually never flagged; the status quo remains that researchers are left to learn about them by soaking up the lore in their particular community. “There is nothing more irritating than publishing a paper that completely disproves every major conclusion of a study, and then years later seeing reviews or other papers cite the original (wrong) study, without the authors being aware that any doubts were ever raised,” says Thomas DeCoursey, a biologist at Rush University Medical Center in Chicago. In 2006, he raised questions about research that had been published in *Nature* two years earlier; the paper was not retracted until November 2010.

Ivan Oransky, executive editor at Reuters Health and co-founder of the blog Retraction Watch, feels that such difficulties are just symptoms of a wider issue with the reward system of academic research: publications are the only way to accrue scientific merit, so they take on a sanctity that academics are reluctant to disrupt with corrections or retractions. If researchers could afford to view scientific output more as a continuous stream, rather than a punctuated series of publications, revisions would carry less of a stigma, he says.

occurred. And that stigma may dissuade honest scientists from doing the right thing. One American researcher who talked to *Nature* about his own early-career retraction said he hoped that his decision would be seen as a badge of honour. But, even years later and with his career established, he still did not want *Nature* to use his name or give any details of the case.

There is no general agreement about how to reduce this stigma. Rennie suggests reserving the retraction mechanism exclusively for misconduct, but that would require the creation of a new term for withdrawals owing to honest mistakes. At the other extreme, Thomas DeCoursey, a biologist at Rush University Medical Center in Chicago, argues for retraction of any paper that publishes results that are not reproducible. “It does not matter whether the error was due to outright fraud, honest mistakes or reasons that simply cannot be determined,” he says.

A better vocabulary for talking about retractions is needed, says Steneck — one acknowledging that retractions are just as often due to mistakes as to misconduct. Also

useful would be a database for classifying retractions. “The risk for the research community is that if it doesn’t take these problems more seriously, then the public — journalists, outsiders — will come in and start to poke at them,” he points out.

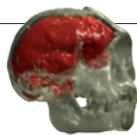
The only near-term solution comes back to transparency. “If journals told readers why a paper was retracted, it wouldn’t matter if one journal retracted papers for misconduct while another retracted for almost anything,” says Zen Faulkes, a biologist at the University of Texas–Pan American in Edinburg, Texas.

Oransky agrees. “I think that what we’re advocating is part of a much larger phenomenon in public life and on the Web right now,” he says. “What scientists should be doing is saying, ‘In the course of what we do are errors, and among us are also people that commit misconduct or fraud. Look how small that number is! And here’s what we’re doing to root that out.’” ■

Richard Van Noorden is an assistant news editor for *Nature* in London.

COMMENT

MEASUREMENT The revolutionary origins of the metric system **p.32**



EVOLUTION Hobbit brains were small but packed a punch **p.34**

BOOKS Margaret Atwood muses on why people prefer dystopian fiction **p.35**

FUNDING Britain to require departments to hold female-friendly awards **p.36**



P. DIEDERICH/NYT/REDUX/EYEVINE

Phosphate is mined to produce fertilizers for crops, but phosphorus leaching into water supplies is an environmental hazard.

A broken biogeochemical cycle

Excess phosphorus is polluting our environment while, ironically, mineable resources of this essential nutrient are limited. **James Elser** and **Elena Bennett** argue that recycling programmes are urgently needed.

To meet our demands for energy, humankind has moved masses of carbon from deep underground into the atmosphere, wreaking havoc with the climate. To meet our demand for food, we have moved large amounts of nitrogen from the atmosphere to fields, rivers and forests, devastating ecosystems. To grow our crops we have interfered with Earth's reserves of a third element — phosphorus — which receives much less press and for which we face the unique problem of having both too much and too little.

Since the middle of the twentieth century,

humanity has quadrupled the environmental flow of phosphorus¹, an essential element for all forms of life. We dug up geological phosphate reserves to produce fertilizers to feed the Green Revolution, creating a largely one-way flow of phosphorus from rocks to farms to lakes and oceans, and dramatically impairing freshwater and coastal marine ecosystems. Globally, oxygen-depleted marine coastal 'dead zones' caused by nutrient-stimulated algal blooms continue to expand. The Gulf of Mexico's dead zone, averaging more than 17,000 square kilometres in recent years, was forecast to

reach record dimensions this year before a tropical storm stirred the waters.

At the same time, concern is growing about how long we can count on cheap supplies of phosphorus for fertilizer: easily mineable deposits of phosphate rock are limited. Unlike nitrogen, phosphorus cannot be pulled from the air and, unlike the carbon in our energy system, there is no known replacement. In 2009, Dana Cordell of the University of Technology in Sydney, Australia, and her colleagues published a 'peak phosphorus' forecast² that predicted maximum production around 2030 — an ►

▶ alarmingly imminent forecast in the light of widespread riots in 2008 sparked by food prices, and a 700% increase in phosphate rock prices from 2007 to 2008.

These issues are not entirely new. In 1938, US President Franklin Roosevelt said it was “high time for the Nation to adopt a national policy for the production and conservation of phosphates for the benefit of this and coming generations”. Astonishingly, such a comprehensive policy never emerged, although in the 1970s, the Tennessee Valley Authority set up the National Fertilizer Development Center to study and expand the production and use of phosphate fertilizers. This was the forerunner of the International Fertilizer Development Center (IFDC), headquartered in Muscle Shoals, Alabama.

New research initiatives are emerging to tackle the two faces of phosphorus, including the Sustainable Phosphorus Initiative at Arizona State University in Tempe, of which one of us (J. E.) is a co-founder. The world continues to face deteriorating water quality, uncertainty about future supplies of phosphorus and uncoordinated institutional frameworks. So we need to move quickly beyond academic discussions to creative policy solutions.

POWER IMBALANCE

Estimates of how much readily accessible phosphorus is left have increased since the wave of concern in 2009. But uncertainties surrounding phosphorus supply remain unreasonably large. Last year, economic geologist Steven Van Kauwenbergh at the IFDC produced an assessment of global phosphorus reserves³, which, by incorporating previously overlooked geological reports from the 1980s, greatly increased the estimated reserves for Morocco and its disputed territory of Western Sahara. This led the US Geological Survey to increase its estimate of accessible global phosphate rock reserves by more than four-fold, from around 15 billion tonnes to around 65 billion tonnes. It is disturbing that these

numbers can change by so much so quickly.

More important than the amount of phosphorus in the ground, is how much it will cost to get it out. Overall, three countries control more than 85% of the known global phosphorus reserves, with Morocco clearly in the driver's seat³ (see ‘Global imbalance’). This concentration of power is far greater than for oil, where the dozen members of the Organization of the Petroleum Exporting Countries control 80% of the world's oil reserves. Such a power imbalance is a potential source of tension, given the political turmoil in northern Africa and the fact that developing-world farmers cannot afford phosphate fertilizers even at today's non-monopoly prices. Major regions of the world have diminishing (United States), few (India) or no (northern Europe) phosphorus reserves of their own. Many of the world's food producers are in danger of becoming completely dependent on trade with Morocco, where press reports have emerged of Dubai-style luxury developments being planned in anticipation of phosphorus windfalls.

The strategic dimensions of phosphorus are beginning to be recognized. In May this year, a workshop sponsored by the US Department of Energy included phosphorus alongside dysprosium, yttrium and other rare earth elements of crucial importance to US national security that face potential supply bottlenecks. Indeed, phosphorus may be included as a ‘strategic material’ in pending US legislation to assess and secure access to sources of key minerals.

RECYCLE, REDUCE

The solutions to these problems lie in recapturing and recycling phosphorus, moving it

from where there is too much to where there is too little, and developing ways to use it more efficiently. Many strategies are simple and readily available, even for poor farmers and developing economies.

Consider the fate of the approximately 17.5 million tonnes of phosphorus mined in 2005, analysed in the paper by Cordell *et al.*². About 14 million tonnes of this were used in fertilizer (much of the rest went into cattle-feed supplements, food preservatives, and the production of detergents and industrial cleaning agents) but only about 3 million tonnes made it to the fork (or chopstick). The largest loss — around 8 million tonnes — was directly from farms through soil leaching and erosion. Much research and effort has already been expended to reduce such losses, including more precise timing and placement of fertilizer along with no-till cultivation, but adoption of these best practices needs to become more widespread. In 2009, the total phosphorus mined had increased to 23 million tonnes, but the general phosphorus pathways and losses have not changed much since the earlier analysis.

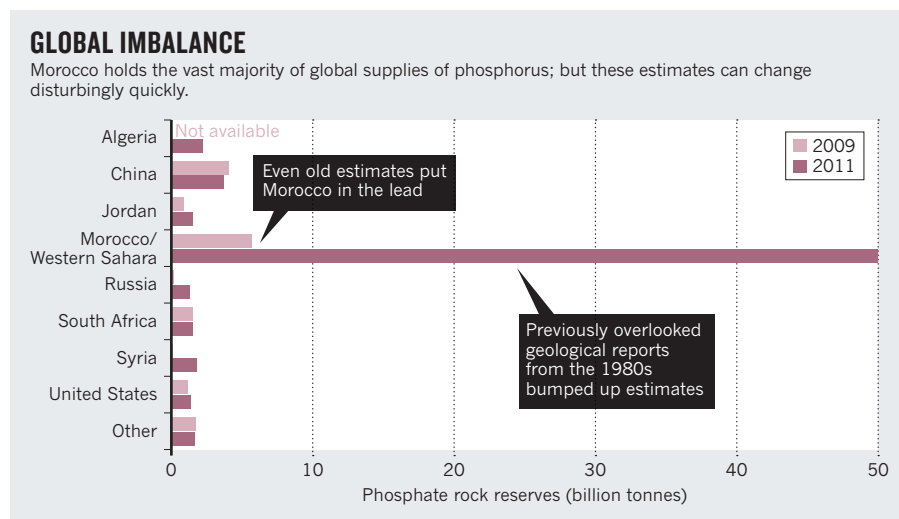
On average, about 30–40% of food produced is spoiled or wasted, and this wastes around 1 million tonnes of phosphorus every year². Producing more food within or closer to cities could reduce waste and facilitate recycling by composting and other approaches.

We can also recycle phosphorus from human waste. Each person excretes about 1.2 grams of phosphorus per day²; trapping all of this globally would produce about 3 million tonnes per year — about 20% of annual worldwide phosphate fertilizer consumption. Currently, only 10% of phosphorus from human waste is returned to agriculture; one method involves extracting ‘struvite’ (magnesium ammonium phosphate) at sewage-treatment plants and processing it into fertilizer pellets.

Urine-separating toilets and latrines can help to capture nutrients for return to the soil as well as improve sanitation in the developing world. Already deployed in Europe, the NoMix toilet captures urine in the front and faeces in the back, diverting the urine for recycling at household, neighbourhood or city scale. Urine-separating latrines are now being installed on a relatively large scale in Durban, South Africa, funded by a grant from the Bill & Melinda Gates Foundation. Another low-cost solution is the Peepoo, a single-use, self-sanitizing, biodegradable bag that captures human excreta and can be used, or even sold, as fertilizer 2–4 weeks later. For the poorest of the poor, it would be a potentially radical transformation if their own ‘waste’ could become a source of income.

According to Cordell *et al.*², more than 7 million tonnes of phosphorus were released into the environment annually in the 2000s through animal manure and excreta,

DATA: USGS



causing major water-quality problems. Even if manure were collected, large livestock farms are now often too far away from arable land for transport of the heavy wastes to be economically viable. Combined bioenergy and waste-trapping technologies can help. For example, bioreactors developed by Bion Environmental Technologies in Crestone, Colorado, are being used at a large dairy operation in Pennsylvania to prevent nutrient run-off to Chesapeake Bay; recovered nutrients are slated for return to farms. However, as with manure, transport costs could limit the scalability of this approach. Market incentives might help to make struvite-recovery systems, such as those developed for municipal wastewater-treatment plants, economic for high-density livestock operations.

Reducing the phosphorus requirement for crops and livestock would make it easier for sources of recycled phosphorus to meet agricultural demand. One way to do this is to encourage people to switch to vegetarianism: producing a vegetarian's diet requires 1 kilogram less phosphorus per year than a meat-eater's. This might be a difficult social change to enact, however.

Researchers have recently engineered some plants to increase their ability to scavenge nutrients, including phosphorus, from soils. One approach, published earlier this year, modified *Arabidopsis*, tomato, rice, alfalfa and cotton to overexpress proton-translocating enzymes called pyrophosphatases, which leads to more elaborate root systems and higher production of leaves, fruit and seeds, among other things⁴. So far, these approaches have been limited to the lab and to test plots.

What about livestock? Promisingly, the gene for bacterial phytase, an enzyme that breaks down the phosphorus-rich compounds called phytates, has been introduced into a line of Yorkshire pigs, creating the Enviropig⁵. These pigs produce phytase in their saliva, which allows them to make use of the otherwise undigestible phytates in feed grains. The transgenic pigs produce up to 75% less phosphorus in their manure than do non-transgenic pigs, and do not need phosphate added to their diet. Twelve years after development, the technology is still working its way through federal approval processes in the United States and Canada.

POLICY MEASURES

Together, these measures would help to cut phosphorus-containing waste, enhancing food security while also reducing the polluting effects of phosphorus run-off. But the gaping institutional vacuum for phosphorus



The Peepoo biodegradable bag captures human excreta and can be used as fertilizer.

governance must also be plugged. Society needs more rigorous, independently verified estimates of pools and fluxes of this critical element as well as reliable ways of estimating the affordability of remaining stocks.

As illustrated by the recent radical adjustment of reserve estimates, we barely know how much phosphorus we have. Current methods to gauge phosphorus reserves rely largely on voluntary provision of proprietary data by private industry and government agencies and, as in the case of Morocco, are often based on relatively old geological assessments. There are other disturbing uncertainties⁶. Estimates of major phosphorus fluxes, such as the loss of phosphorus from agricultural lands, span a three- to fivefold range; others, such as the global return of phosphorus from harvested crops to farms, are essentially unknown.

Attempting to bridge these gaps, new networks of sustainability scientists have come together — as yet loosely organized and lightly funded. Among the first was the Global Phosphorus Research Initiative (now the Global Phosphorus Network), which produced some of the first estimates of potential time frames for phosphorus scarcity⁷. In 2010, the Global Transdisciplinary Processes for Sustainable Phosphorus Management consortium emerged, to connect scientists, industry, business, and government groups at each point of the phosphorus supply and use chain. The Sustainable Phosphorus Initiative sponsored a summit in February 2011 which produced the Phoenix Phosphorus Declaration: a consensus of more than 100 scientists, engineers, architects, designers, farmers, entrepreneurs, artists and communicators on the urgency and opportunities associated with achieving phosphorus sustainability.

Sadly, the message hasn't yet sunk in where it counts. The 2009 United Nations Food and Agriculture Report and the 2010 report of the US National Research Council Committee

on Twenty-First Century Systems Agriculture breathe hardly a word about fertilizer supplies, prices and access, instead focusing on the impacts of fertilizer run-off on water quality, and generally emphasizing the effects of excess nitrogen rather than phosphorus. More promisingly, both Sweden and Germany are implementing ambitious directives to recycle up to 60% of wastewater phosphorus, with half of it to be returned to farms and the rest to pastures or forest plantations.

To move further and faster, we call for the establishment of a comprehensive network of national and international science and policy research centres for nutrient sustainability, which should also tackle nitrogen (which has a 'too much' problem⁷) and potassium (which may have similar geopolitical issues of 'too little'). Such centres should pursue research both on fundamental biogeochemical processes in agriculture and on possible policy actions, working closely with practitioners and policy-makers.

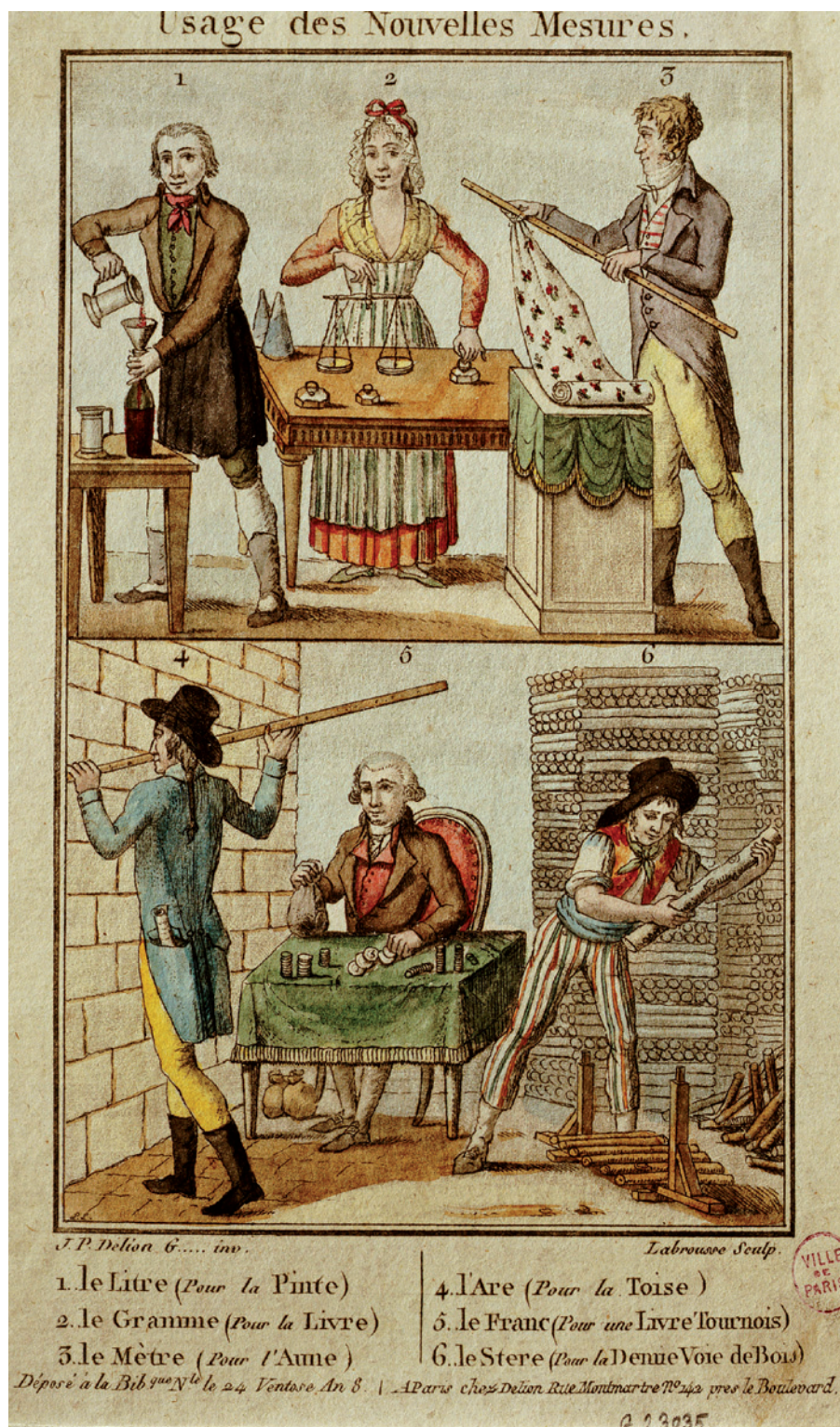
One idea is to create phosphorus-emission markets, similar to carbon markets. Some jurisdictions, including 13 US states and some Australian states, are considering these now, primarily to protect water quality. For example, in an arrangement under discussion in Maryland, an advanced wastewater-treatment plant that surpasses federal guidelines for nutrient releases could, through a private broker, sell permissions to release nutrients to other municipalities whose facilities cannot currently meet the targets.

Another, as yet unconsidered, idea might be the creation of national or international strategic phosphorus reserves, similar to the petroleum reserve, to stabilize commodity prices.

As Roosevelt said in 1938: "I cannot over-emphasize the importance of Phosphorus not only to agriculture and soil conservation but also to the physical health and economic security of the people of the nation." Nearly 75 years later, it is time to find long-term solutions for one of life's essential elements. ■

James Elser is at the School of Life Sciences, Arizona State University, Tempe, Arizona, USA. **Elena Bennett** is at the Department of Natural Resource Sciences and McGill School of Environment, McGill University, Montreal, Quebec, Canada.
e-mail: j.elser@asu.edu

1. Falkowski, P. *et al. Science* **290**, 291–296 (2000).
2. Cordell, D., Drangert, J.-O. & White, S. *Global Environ. Change* **19**, 292–305 (2009).
3. Van Kauwenbergh, S. J. *World Phosphate Rock Reserves and Resources Technical Bulletin IFDC-T-75* (IFDC, 2010).
4. Gaxiola, R., Edwards, M. & Elser, J. J. *Chemosphere* **84**, 840–845 (2011).
5. Golovan, S. P. *et al. Nature Biotechnol.* **19**, 741–745 (2001).
6. Childers, D. L., Corman, J., Edwards, M. & Elser, J. J. *BioScience* **61**, 117–124 (2011).
7. Sutton, M. A. *et al. Nature* **472**, 159–161 (2011).



The metric system was invented in France in the early 1790s but took 200 years to become dominant.

HISTORY

Scaling up

Andrew Robinson applauds a chronicle of metrication that balances physics with philosophy.

The metric system of measurement was invented by French scientists in the early 1790s and imposed on the populace by the leaders of the French Revolution. For about a year, the revolutionaries even attempted to introduce a decimal clock, with each day divided into ten hours, each hour into 100 minutes and each minute into 100 seconds. Napoleon Bonaparte later congratulated the scientists: “Conquests will come and go, but this work will endure.”

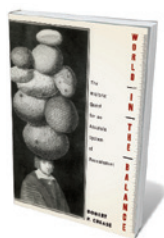
Yet Napoleon himself refused to use metric units. In 1812, he ordered their official withdrawal. And, after falling from power in 1815, he attacked the system as a “stumbling block” to progress. Most of the French people also rebelled. Not until 1840 did the government again dare to make use of metric measures obligatory.

From France, the metric system spread haltingly around the globe, as *World in the Balance* documents. Robert Crease, a philosopher at Stony Brook University in New York, relates the history of measurement from ancient China to current debates over defining the kilogram by reference to physical constants. His respect for both the philosophical and physical aspects of measurement adds tension to his account.

Crease explains how growing industrialization and mechanization promoted the metre and suchlike. The advantages were also highlighted by events such as the Great Exhibition in London in 1851, which revealed the incompatibility of different national measurement systems, and by cheerleading editorials in *Nature* from the 1870s.

In 1875, representatives of 17 nations and empires met in Paris to sign the Convention of the Metre, “desiring international uniformity and precision in standards of weight and measure” and establishing the International Bureau of Weights and Measures (BIPM) at Sèvres. The United Kingdom and its colonies signed only in 1884, but in practice continued to use imperial measures for nearly another century. The Soviet Union officially went metric in the 1920s; Japan in the 1950s. The BIPM introduced the International System of Units (SI units, from the French) in 1960. Not until the final decades of the twentieth century did the metric system become dominant around the world.

Only the United States, Myanmar (Burma) and Liberia have yet to enact legislation to metricate. Even so, scientists and many of the public in the United States use the metric system daily. The third US president, Thomas Jefferson — Francophile, fanatical quantifier and admirer of the metric system — tried to convert his countrymen, but eventually gave up. “Shall we mould our citizens to the law, or the law to our citizens?” he wrote in an 1817 letter to John Quincy Adams, a future president who in 1821 reported to the US government on the feasibility of metrication.



World in the Balance: The Historic Quest for an Absolute System of Measurement

ROBERT P. CREASE
W. W. Norton: 2011.
288 pp. £20

Adams favoured the first option; but his report to Congress plumped for the second, and recommended retaining the existing system. Metrication, said Adams, would “affect the well-being of every man, woman and child”. The change would be too radical. The physicist and polymath Thomas Young, surprisingly omitted by Crease, gave the same advice to the British Parliament in the 1820s.

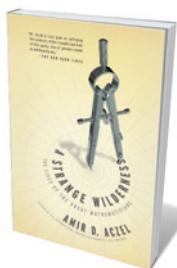
The most original section of the book concerns ancient China. Writing with the help of Chinese metrologists, Crease describes how the 12-note harmonic scale used in ritualized music helped to define the measurement system used at the imperial court. The ritual scale, known as the *lǚlǚ*, was supposedly devised by China's first emperor, Huang Di, in the third millennium BC. He sent a minister to the mountains to procure bamboo of a species revered for its regularity in length and thickness. From a piece 3.9 *cun* in length — the *cun* was the width of a thumb knuckle, or one-tenth of a *chi*, the ‘Chinese foot’ — Huang Di made a one-note flute, whose pitch became the lowest in the scale, known as the *huangzhong*. Eleven more bamboo flutes created the *lǚlǚ*.

Whatever the truth of this legend — the evidence suggests that the 12-note scale was actually introduced much later, some time before 400 BC — 12 pitch regulators were made in cast metal; the lengths were specified by regulation in *chi*, linking the basic unit of length with musical pitch. This system endured for more than 2,000 years and was not replaced until the 1920s, with the adoption of the metric system. In 1984, the country defined the *chi* to be one-third of a metre.

As a physicist, Crease is drawn to the drive for quantification, uniformity and precision in measurement — the main concern of his book. As a philosopher, he understands that there is more to quantification than these scientific virtues. The value of education and of scientific research, for example, cannot be measured wholly by examination results and citation indices. Nor can the fitting of clothes be entirely mechanized, as Crease concludes, after trying out various body scanners used by US retailers. Science cannot proceed without precise measurement, yet successful measuring systems cannot be divorced from everyday human dimensions. ■

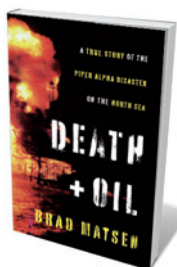
Andrew Robinson is the author of *The Story of Measurement*.
e-mail: andrew.robinson33@virgin.net

Books in brief



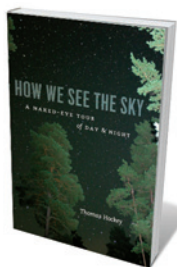
A Strange Wilderness: The Lives of the Great Mathematicians
Amir D. Aczel STERLING 304 pp. \$24.95 (2011)

A poet-mystic; a swordsman clad in green taffeta; a 12-year-old who mastered ancient Greek. Omar Khayyam, René Descartes and Gottfried Leibniz are just three of the mathematical greats in Amir Aczel's trot through theorems and the lives behind them. Aczel, author of *Fermat's Last Theorem* (1996), begins with the Greeks; ponders the geniuses of India, Arabia and China; frolics in the hotbed of the Italian Renaissance; examines the founders of calculus and the wunderkinder of the Napoleonic age; and skids to a halt with Alexander Grothendieck, who learnt maths in a Nazi internment camp.



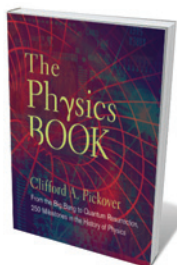
Death and Oil: A True Story of the Piper Alpha Disaster on the North Sea
Brad Matsen PANTHEON 224 pp. \$25.95 (2011)

More than two decades before the Deepwater Horizon oil spill, the Piper Alpha oil rig exploded in the North Sea, killing 162 men. Writer Brad Matsen has packed in two years of research, during which he has interviewed survivors, managers, rescue teams and government officials. Matsen is thorough in laying out the scientific, technological, industrial and political context. This is a deftly told tale of human error, technological glitches and corporate reluctance that highlights the high cost of our thirst for crude.



How We See the Sky: A Naked-Eye Tour of Day and Night

Thomas Hockey UNIVERSITY OF CHICAGO PRESS 224 pp. \$60 (2011)
Images of the Horsehead Nebula from the Hubble Space Telescope are more familiar to most of us than the sight of the sky above our heads. So argues astronomer Thomas Hockey, who urges us to gaze unaided at the Universe. Starting with a scan of the horizon, Hockey takes us through the science as well as a host of cultural references, from Pink Floyd to the Pyramids of Giza in Egypt. He explores the astronomical sky, the 88 constellations and the Milky Way; orientation through azimuth to zenith; lunar and solar motion, solstices and eclipses. A heavenly and often humorous journey.



The Physics Book: From the Big Bang to Quantum Resurrection, 250 Milestones in the History of Physics

Clifford A. Pickover STERLING 528 pp. \$29.95 (2011)
Molecular biophysicist and inventor Clifford Pickover follows his 2009 volume *The Math Book* with this energizing look at 250 discoveries in physics. Bookended by the Big Bang and the ‘quantum resurrection’, the landmark events run from Archimedes' burning mirrors, Isaac Newton's prism, the Higgs boson and the Doppler effect to dark energy, Wolfgang Pauli's exclusion principle and rogue waves. Luminaries from Archimedes to Fritz Zwicky get their due, and it is gorgeously illustrated throughout.



Great Discoveries in Medicine

Edited by William Bynum and Helen Bynum THAMES & HUDSON 304 pp. £24.95 (2011)
Dazzling images adorn this crisply written chronicle of ‘eureka’ moments in medicine, covering our emergent knowledge of the body, diseases, drugs and surgery. A drawing of a Caesarean section in Hermann Friedrich Kilian's nineteenth-century *Geburtschülfflicher Atlas* has the delicacy of Flemish Renaissance art. Other marvels include the first X-ray (of Mrs Röntgen's ringed hand), a photograph of serotonin crystals and computer-generated images of viruses.

PALAEOANTHROPOLOGY

Craniums with clout

A look at two early human fossils reveals the prejudices in ideas about human evolution, finds **Henry Gee**.

We have all seen the canonical parade of apes, each one becoming more human. We know that, as a depiction of evolution, this line-up is tosh. Yet we cling to it. Ideas of what human evolution ought to have been like still colour our debates.

Palaeoanthropologist Dean Falk debunks some modern myths in her brilliant book, *The Fossil Chronicles*, by comparing the case histories of two famous fossils. A career spent teasing meaning from the brain casts of fossil hominins (creatures more closely related to *Homo sapiens* than to chimpanzees) has led Falk into the debate on the cognitive abilities of *Homo floresiensis*. This dwarfed hominin — nicknamed the Hobbit — lived on the Indonesian island of Flores between approximately 95,000 and 14,000 years ago, and was discovered in 2003 (see *Nature* 431, 1055–1061; 2004). Falk also describes Raymond Dart's 1924 discovery in South Africa of a juvenile skull of *Australopithecus africanus*, the Man-Ape of South Africa, and locates an unpublished manuscript by Dart on the find that chimes with her views.

"Brains might be small, but they can still pack a punch."

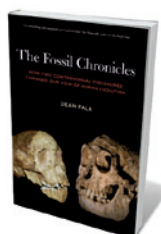
Almost every time someone claims to have found a new species of hominin, someone else refutes it. The species is said to be either a member of *Homo sapiens*, but pathological, or an ape. Brickbats of the first kind were levelled recently at *H. floresiensis* — that it wasn't a genuine species, but a modern human suffering from one of several kinds of microcephaly or from cretinism. But they had also been aimed at Neanderthal Man, discovered back in 1856, and thought by some to be the remains of a Mongolian Cossack from the Napoleonic wars. Accusations of apishness were aimed at *A. africanus*, described by Dart in these pages in 1925; and at *Sahelanthropus tchadensis*, nicknamed Toumaï, a very primitive putative hominin from Chad, discovered in 2001.

Dart's original paper on *A. africanus* was, it is true, long on waffle and short on substance. But the reason that this small-brained, possibly erect-walking creature took two decades to be accepted as a hominin was

that researchers were in thrall to the idea that the expansion of the human brain came first, before the adoption of a fully erect gait. This preconception was supported by the discovery of the large-brained, ape-jawed Piltdown Man in 1912. The fact that it took 40 years to expose Piltdown as a fraud is a mark of how deeply rooted such prejudices can be.

Falk describes her work refuting the idea that the small brain of the Hobbit implies the creature might have had a congenital disorder of brain growth. She shows that its brain most resembled that of *Homo erectus*, another antique hominin, and was developed in areas associated with cognitive abilities that would have supported making the simple tools with which its fossils are associated. Yet the Hobbit has a closer resemblance in its general anatomy to *Australopithecus*, suggesting — again contrary to preconception — that hominins emerged from Africa much earlier than thought.

The best parts of the book are those in which Falk traces the history of Dart. The Australian anatomist was exiled to South Africa by his mentor at University College London, Grafton Elliot Smith. In Africa, chance threw in his way the brain cast and skull of a juvenile hominin: a creature he named *A. africanus*. After publishing his *Nature* paper, Dart was critically mauled by the London establishment — notably the 'Piltdown



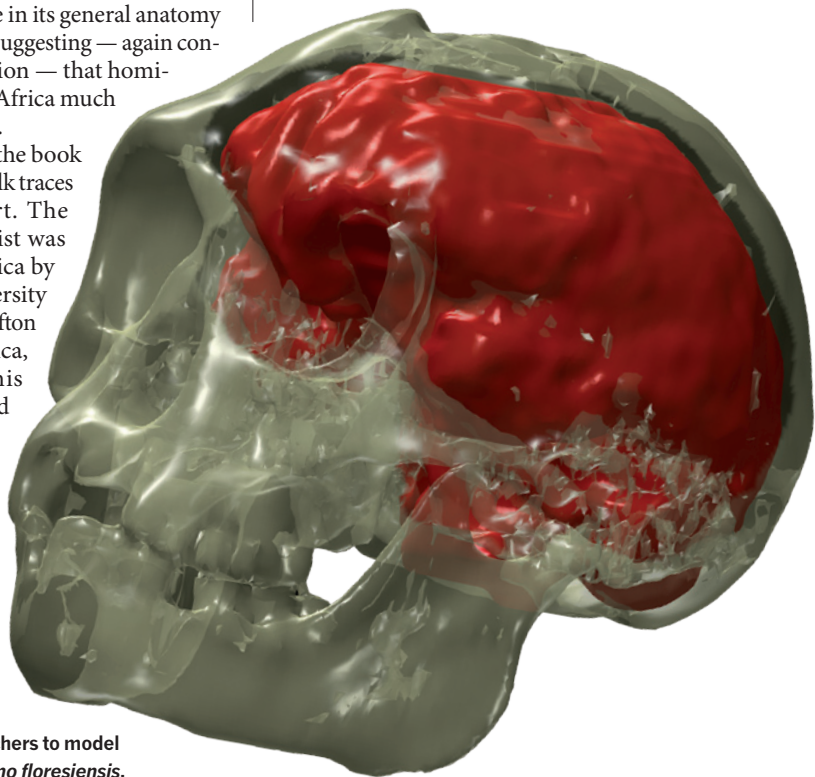
The Fossil Chronicles: How Two Controversial Discoveries Changed Our View of Human Evolution
DEAN FALK
University of California Press: 2011. 280 pp.
\$34.95

Committee' who believed in the fake fossil — and he almost deserted palaeoanthropology, devoting his energies to building up capacity at the then-fledgling University of the Witwatersrand in Johannesburg.

Almost, but not quite. Falk's investigation of Dart's papers at Witwatersrand has brought to light a monograph on *A. africanus* that Dart never published. In 1929 he sent it to Elliot Smith to submit to the Royal Society in London, but it was rejected, presumably on the basis of reports by the Piltdown Committee. Falk reveals that Dart had come to similar conclusions about the cognitive capacity of *A. africanus* as she has with *H. floresiensis*, providing circumstantial evidence for her link between *Australopithecus* and the Hobbit, and for an earlier African diaspora. Brains might be small, but they can still pack a punch.

Falk's book is worth reading just for the unearthing of this otherwise lost manuscript, vital to the history of palaeoanthropology. That it sparkles with scholarship and wit is icing on the cake. ■

Henry Gee is a Senior Editor of *Nature*.



Scans enabled researchers to model the brain shape of *Homo floresiensis*.

➔ **NATURE.COM**
For more on *Homo floresiensis* skull scans:
go.nature.com/xzbbgc

K. C. ARMSTRONG/CORBIS



Q&A Margaret Atwood

Speculative realist

Novelist Margaret Atwood's essay collection *In Other Worlds: SF and the Human Imagination*, published this month, is a companion piece to her dystopian fictional world of global warming and engineered plagues. The Canadian author discusses where she gets her science, and her concerns for the future.

Does science run in your family?

My father was an entomologist — he studied sawflies, budworms and insects that eat trees, so as a child I spent a lot of time in the forest. My brother is a neuroscientist who studies synapses, one nephew is a physicist studying the composition of the Universe, another is a materials engineer studying crystal structure. My grades were a bit better in science than in English, so I easily could have become a biologist: I'd probably be cloning potatoes now, making them glow in the dark. But I started writing instead.

You say in your new book that your novels are not science fiction, but speculative fiction. What's the difference?

It is hard to draw that line. A lot of what is labelled science fiction has nothing to do with science. It tends to be something that doesn't fit into any other genre, so it is all put in the same box. But to me there is a difference between a science-fiction novel such as Ursula LeGuin's *The Left Hand of Darkness* — which contains things that are very unlikely to happen, or impossible — and a speculative novel such as George Orwell's *1984*, which really could happen. My books are more like the latter — I don't write about Planet X.

In Other Worlds: SF and the Human Imagination

MARGARET ATWOOD
Nan A. Talese/Virago:
2011. 272 pp.
\$24.95/£17.99

You also note that we're preoccupied today with dystopias. Why is that?

We're not feeling very hopeful about our future. In the nineteenth century, everybody thought they had a bright idea that would make life better. We wrote about utopias and model communities. The future was seen as a place of infinite advance. Then came the two World Wars and a number of totalitarian societies that came in on a utopian ticket. The Soviet Union promised wonderful things and put on a good show, but meanwhile Stalin was starving Ukraine and butchering millions of people. We remember those experiences and know too much about them. It has become less and less possible to write a utopia that isn't some form of *Stepford Wives* or *Brave New World*.

What sort of future do you imagine in your books *Oryx and Crake* and *The Year of the Flood*?

Genetic engineering is commonplace. A scientist named Crake designs a race of improved humans that are better adapted to their environment. They don't have to wear clothes because they've got built-in sun block and insect repellent. They'll never have to farm because they eat leaves. They're all beautiful and mate seasonally, so there's

no sexual jealousy. And they will drop dead at the age of 35, so they won't have age-related illnesses. To make room for them, Crake arranges to eliminate everybody else with a bioengineered epidemic. Having fun yet? However, not everybody is eliminated. *Oryx and Crake* is told from the point of view of one survivor. In *The Year of the Flood*, which tells a parallel story, we find that a few other people have also survived because they took precautions.

How do you keep track of science?

A number of scientists follow me on Twitter. They pass along reports of advances such as transplanting human brain cells into animals, or making meat in the lab, or creating a new gene. Some of the things I wrote about in *Oryx and Crake* hadn't actually happened then, although you could see them coming and they have been done since. Other things that people thought I'd made up, like the goat-spider mix and the light-up rabbit, were already real.

In *Other Worlds* cautions that, given the risks of biotechnology and cryogenics, "we should leave well enough alone". Why?

Humans will play with their toys until something blows up. Once you let it out of the box, it is hard to put it back in. We now have the ability to create human-specific diseases to which nobody has any immunity and deploy them simultaneously all over the world. Cryogenics, on the other hand, is a nonstarter: you get your head frozen, the money runs out, your relatives die, and you're cat food.

Why does science scare some people?

Science is attractive to those who like solving puzzles. But it is not so appealing for people who want to be cuddled (or even reprimanded), who want to feel that things make sense, or that somebody's looking after them. Scientists do not offer certainty, and they do not offer a universe that is centred around humans. Religions offer a world view in which you are important.

Does the future worry you?

I'm past the age when things scare me. But if I were younger, I would be looking down the line with some apprehension. A world with more than 9 billion people is not going to be very habitable. We've already used 90% of the fish in the sea. Global warming will make it worse: more droughts, more extreme weather and limited harvests. People think they will fix the problem with technology, but famine may fix it for us. Either way it will be a pretty miserable life. The infinite inventiveness of humans sometimes makes me feel hopeful, but we're just as capable of inventing horrible things as good things. ■

INTERVIEW BY JASCHA HOFFMAN

➔ **NATURE.COM**
For author Tom Wolfe's take on science:
go.nature.com/mepdrq

Correspondence

Reduce drug waste in the environment

Environmental contamination by pharmaceuticals is reaching alarming levels (see, for example, *Nature* **476**, 265; 2011) and is set to rise. New partnerships between drug companies, the public-health sector and those who deliver environmental sustainability are urgently needed to tackle the issue.

Low-cost pharmaceuticals are increasingly accessible to the global population, which is predicted to exceed 8 billion by 2050. Rising drug use is also driven by ageing populations. Widely used preventative medication — such as statins and anti-hypertensives — and cheap generic drugs add to the problem. The UK Office of National Statistics predicts that the country's medicine usage will more than double by 2050.

Agricultural soils and rivers are contaminated with a range of pharmaceuticals, including antibiotics, antidepressants, analgesics and cancer-chemotherapy agents (see go.nature.com/lr2vfy).

The effects are already evident: they include the feminization of fish by residues of the contraceptive pill, and the deaths of millions of vultures on the Indian subcontinent following ingestion of the anti-inflammatory drug diclofenac. Antibiotic overuse has led to the emergence of resistant pathogenic bacteria in the wider environment, and not just in medical settings.

Current practices remain unchanged. However, attempts are being made to provoke action. The European Environment Agency has recommended that improvements be made in pharmaceutical-waste management, and that more guidance be provided for the public and for policy-makers. The UK Royal Commission on Environmental Pollution in March highlighted links between demographic change



and pharmaceutical releases, and the UK government's Advisory Committee on Hazardous Substances will conduct an investigation.

Michael Depledge *European Centre for Environment and Human Health, University of Exeter, UK.*
michael.depledge@pms.ac.uk

Bridging the gender gap in UK science

Sally Davies, the Chief Medical Officer for England, has broken new ground for gender equality in the sciences in a letter to the UK Medical Schools Council on 29 July. She will make it a requirement for academic departments applying for funding from the English National Institute for Health Research to hold the silver award of the Athena SWAN (for 'scientific women's academic network') Charter. We urge other funding bodies, including the UK research councils and the Royal Society, to follow suit.

The charter recognizes good employment practice for women in UK science, engineering and

technology (SET). It is supported by the Equality Challenge Unit and the UK Resource Centre for Women in SET.

The charter invites applications from UK universities and university-linked research institutes and departments to apply for bronze, silver or gold awards. These awards promote career development as well as gender equality. For example, they encourage improved mentoring and guide parents in how to partition their time between academic and family life.

Departments can only apply for a silver award if their university already holds a bronze award. If a department does not get its silver award immediately, Athena SWAN advisers can recommend improvements. Davies's silver-award requirement will come into play for the next round of funding in four years' time — a practical measure that gives universities and departments time to get their bronze and silver awards.

The Athena SWAN Charter currently recognizes 35 bronze universities, 11 bronze departments, 40 silver

departments and one gold department (the University of York's chemistry department). The next deadline for submissions is 30 November. **Athene Donald** *University of Cambridge, UK.*

Paul H. Harvey, Angela R. McLean *University of Oxford, UK.*
paul.harvey@zoo.ox.ac.uk
Competing interests declared. See <http://dx.doi.org/10.1038/478036b>.

Cloning advance calls for careful regulation

In this issue, Scott Noggle and colleagues describe the generation of human pluripotent stem cells using somatic cells and human oocytes, a technique that bypasses ethical concerns about exploiting fertilized embryos for their medical potential (*Nature* **478**, 70–75; 2011).

The cell lines were produced at the New York Stem Cell Foundation using private funds, in accordance with the Empire State Stem Cell Board's policy of compensating egg providers for research. Unfortunately, many scientists will not have access to these cells, owing to regulations that prevent the publicly funded use of stem cells derived from research embryos and compensated egg donors. These policies — including those in place in California and at the US National Institutes of Health — are well intentioned, but possibly misguided.

To generate their stem-cell lines, Noggle *et al.* use human oocytes in a new twist to the cloning technique known as somatic-cell nuclear transfer (SCNT). Societal fears about reproductive cloning should not force knee-jerk legislation to ban all forms of human SCNT. The use of this technique in research has a clearly regulated goal: to provide patient-specific stem-cell lines to help treat human disease, just like the almost universally supported research involving induced pluripotent

stem cells, which are derived artificially from somatic cells.

Regulatory policies for SCNT in different states and countries must be in agreement, and should adhere to the ethical guidelines for oocyte procurement issued by the International Society for Stem Cell Research. This will help to ensure that SCNT research proceeds with the requisite oversight and fair recruitment and compensation practices for oocyte providers (see also *Nature* **442**, 629–630; 2006).

Insoo Hyun, Paul Tesar *Case Western Reserve University, Cleveland, Ohio, USA.*
insoo.hyun@case.edu
Competing financial interests declared. See <http://dx.doi.org/10.1038/478036c>.

Energy should form its own discipline

The international energy system needs an overhaul. The sector is multidisciplinary: it must serve modern civilization without compromising economic opportunity, undermining national security or impinging on the environment. Yet innovation today prioritizes improvements to discrete technologies and progress in single disciplines rather than rebuilding the whole system. A more joined-up approach is needed, beginning with education.

Retooling the system will require a range of experts who understand new technologies and can translate them to the public, while considering the economic drivers necessary for their adoption.

In the United States, for example, the educational framework for undergraduates does not always keep pace with advances in science, engineering and innovation. Even though energy is a leading international priority, it lacks definition in universities, where it is largely perceived as a professional pursuit, or as a subset of fields such as petroleum engineering. Often, students are exposed only to glimpses of the sector and do not acquire an integrated,

systems-level perspective.

Whereas institutions such as Duke University in Durham, North Carolina, the University of Texas at Austin and the University of British Columbia in Vancouver, Canada, have created programmes to address the changing energy landscape, none offers an interdisciplinary energy-focused degree at undergraduate and graduate levels.

We propose that large energy departments should be set up at universities worldwide to tie seemingly disparate fields of knowledge together. Graduates could move between disciplines to promote ideas and work towards practical solutions. By fostering an open dialogue between specialists, this nascent labour force would then be well equipped to navigate through all of the technical, political and social issues related to energy.

Sheril R. Kirshenbaum, Michael E. Webber *University of Texas at Austin, Texas, USA.*
sheril.kirshenbaum@mail.utexas.edu

Giant dam threatens Brazilian rainforest

Brazil's rainforest is under further threat from plans to build a giant hydroelectric dam on the Xingu River, a tributary of the Amazon River in Pará state. Plans for the dam, known as Belo Monte, have been approved by the environment agency. These come on top of pending changes to the Brazilian Forest Code that could allow deforestation of up to 20 million hectares of rainforest (*Nature* **476**, 259–260; 2011).

The US\$17-billion dam, together with four planned upstream dams, will have a combined hydroelectric potential of 21,600 megawatts. Leaders of the Brazilian energy sector argue that the dams could help to preserve the Amazon. But their construction will flood vast areas of tropical rainforest, jeopardizing ecosystem functions and species survival, increasing greenhouse-gas emissions and displacing tens of thousands of

forest peoples.

Brazil is a world leader in clean-energy production. However, the dams will release into the atmosphere enormous quantities of methane — a greenhouse gas that is 25 times more potent than carbon dioxide.

Much of the electricity generated by Belo Monte is likely to be used in the production of aluminium ingots for export (see go.nature.com/latlx3), making the environmental and social impact of the dam's construction even harder to justify.

Brazil must strive to control deforestation more effectively by strengthening its forest laws and consolidating the United Nations' REDD policy (for 'reduced emissions from deforestation and forest degradation'). Otherwise, the steady destruction of the country's tropical rainforest will have consequences well beyond its borders.

Alison G. Nazareno *Federal University of Santa Catarina, Florianópolis, Brazil.*
alison_nazareno@yahoo.com.br
Thomas E. Lovejoy *Heinz Center for Science, Economics and the Environment, Washington DC, USA.*

Pilot scheme for misconduct database

Researchers, journal editors and scientific institutions should work together to improve communication about misconduct cases. Although published retractions are logged by PubMed and other databases, and by blogs such as Retraction Watch (<http://retractionwatch.wordpress.com>), the scientific community needs a way to identify flawed articles that have not been formally retracted but have been assessed as containing falsified data or having ethical problems (see, for example, *Nature* **476**, 263–264; 2011).

To this end, we have piloted an open database of publications for which misconduct has been established by committees (such as offices of research integrity within research institutions). The database is collaborative

and is coupled to an online platform on which scientific integrity can be openly and constructively debated (see www.scientificredcards.org and T. Fluttre *et al.* *Eur. Sci. Ed.* **36**, 51–52; 2010).

The website focuses on the publications and not the authors, to avoid 'naming and shaming'. It has been legally validated by the French National Commission on Informatics and Liberty, so that such information can be made public while respecting privacy laws.

To expand this initiative, the legal implications would have to be considered. It would need to be endorsed by the research community, which would cooperate to maintain and moderate it. Extensive publicity would be essential to ensure that the facility is used effectively.

Our pilot project offers a route to reinforcing society's trust in science. Creating a public library of misconduct through a collaborative web platform is a timely, transparent and efficient way for the research community to communicate about possible scientific impropriety.

Timothée Fluttre *University of Chicago, Illinois, USA.*
Thomas Julou *Ecole Normale Supérieure, Paris, France.*
Livio Riboli-Sasco *Paris Descartes University, France.*
Claire Ribault *Université Paris Diderot, France.*
contact@scientificredcards.org

Discovery inspires those seeking tenure

I strongly disagree with David Helfand's view that tenure is the "social filter" that selects for professors who are "most attracted to lifetime security" (*Nature* **477**, 158–159; 2011).

For most young scientists attempting to scale the ladder of modern academia, I wager that the promise of scientific and intellectual discovery and the chance to inspire younger people are the real incentives.

Barbara-Ann Lewis *Northwestern University, Evanston, Illinois, USA.*
b-lewis@northwestern.edu

FORUM Stem cells

Triple genomes go far

A technique called somatic-cell nuclear transfer has been applied to human oocytes, resulting in the generation of personalized stem cells, albeit genetically abnormal ones. Two experts discuss the biomedical significance of this work and the ethical issues surrounding the use of human oocytes in research. [SEE ARTICLE P.70](#)

THE PAPER IN BRIEF

- Somatic-cell nuclear transfer (SCNT) involves replacing the genome of an oocyte with that of an adult cell.
- Once the 'reconstructed' cell has developed into a blastocyst (a mass of 70–100 cells), stem-cell lines can be derived.
- Human oocytes manipulated by SCNT do not develop to the blastocyst stage.
- To overcome this problem, Noggle *et al.*¹ (page 70) added the nucleus of a differentiated adult cell to an oocyte that still contained its nucleus (Fig. 1).
- This allowed growth to the blastocyst stage, but, undesirably, the resulting cells

had three genome copies — one from the haploid oocyte and two from the diploid differentiated cell.

- Nonetheless, the adult genome copies reverted to gene-expression programs characteristic of embryonic stem cells.
- Moreover, the stem cells isolated from the blastocysts could differentiate into cells of all three germ layers, from which all the tissues and organs of the body develop.
- Noggle and colleagues paid women for their oocytes.
- There are significant legal and social concerns about obtaining human oocytes for research and even therapy.

changes, reprogramming can leave vestiges of the original differentiated (somatic) cell's identity — known as epigenetic memory — through faulty remodelling of chemical modifications on DNA and its associated proteins⁴.

Although it is premature to conclude that these foibles of iPS cells pose insurmountable risks, comparative studies of mouse stem cells suggest that SCNT may be more effective than forced expression of transcription factors in reprogramming cells to a pristine state of pluripotency and erasing epigenetic memory^{5,6}. But until now, discussions of the relative merits of human SCNT-ES cells and iPS cells have been purely theoretical: although successful in non-human primates⁷, the generation of ES cells through SCNT has thus far failed in humans, largely because human oocytes have not been readily available for research.

With the advantage of ready access to a large number (270) of donor oocytes, Noggle *et al.*¹ performed a rigorous exploration of SCNT and identified obstacles to the generation of normal human blastocysts by this technique. The researchers found that products of SCNT in humans stop dividing at the 6–10-cell stage, because removal of the oocyte genome apparently depletes the cell of factors that

Imperfect yet striking

GEORGE Q. DALEY

Noggle and colleagues' study¹ is noteworthy for generating the first — albeit genetically abnormal — human pluripotent stem cells through oocyte-mediated reprogramming and for highlighting major technical barriers to SCNT using human eggs.

Since the first isolation of human embryonic stem (ES) cells in 1998, a compelling strategy for the future envisaged exploiting SCNT to generate personalized embryonic stem cells. The aim has been to reprogram a patient's differentiated cells to pluripotency — the potential to produce any tissue — and then to coax the resulting SCNT-ES cells to develop into disease-relevant cells, either for mechanistic studies or for combined gene and cell therapy². Realistically, however, SCNT is a cumbersome process that cannot be readily scaled to allow widespread therapeutic use.

One breakthrough was the discovery that skin cells can be reprogrammed to a pluripotent state by enforced expression of only four transcription factors linked to pluripotency in ES cells³. The resulting induced pluripotent stem (iPS) cells, whether mouse or human,

are functionally comparable to ES cells and provide an alternative to SCNT for generating personalized stem cells for disease modelling or cell-based therapies free of the problems of rejection.

Despite enthusiasm for iPS cells, however, closer scrutiny of their genetic integrity and differentiation behaviour has revealed subtle yet potentially significant differences from ES cells. As well as provoking rogue genetic

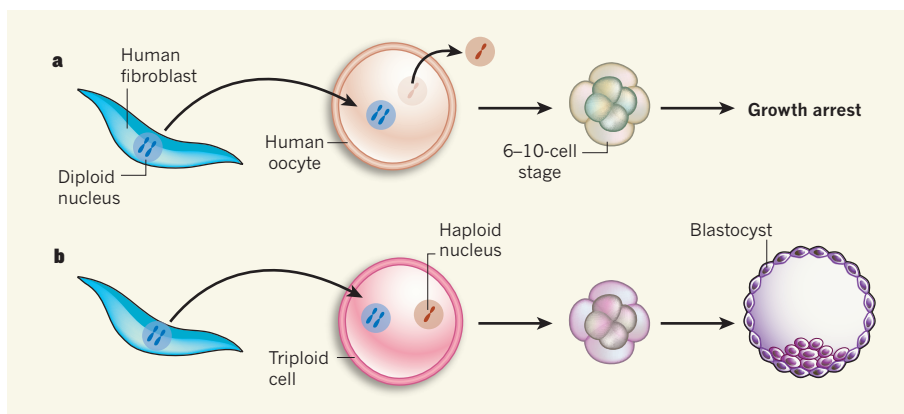


Figure 1 | Three genomes are better than two? **a**, Typically, when the diploid nucleus of a differentiated adult human cell such as a skin fibroblast is transferred into a nucleus-free human oocyte, the resulting cell does not develop to the desired blastocyst stage. **b**, Noggle and colleagues¹ show that leaving the haploid nucleus of the oocyte behind results in the generation of triploid cells that develop to the blastocyst stage. The authors isolated stem cells from these blastocysts (not shown) and found that the derived cells could differentiate into various cell types.

are essential for embryonic cell division or expression of genes from the somatic genome. Frustratingly, they could not overcome this cleavage arrest unless they left the oocyte genome in place; the cells they derived from the resulting blastocysts were therefore triploid somatic–oocyte pluripotent stem cells. Nonetheless, the authors' sophisticated analysis revealed that the transplanted genome was fully reprogrammed, with no signs of epigenetic memory. Thus, although falling short of its ultimate goal, the paper¹ stands as a stepping stone towards success, and raises the provocative question of how human SCNT-ES cells might perform relative to iPS cells.

George Q. Daley is in the Stem Cell Transplantation Program, Division of Pediatric Hematology/Oncology, Howard Hughes Medical Institute, Children's Hospital Boston, Boston, Massachusetts 02115, USA. e-mail: george.daley@childrens.harvard.edu

Persons versus things

JAN HELGE SOLBAKK

What are oocytes? What is their nature? What conceptual labels should be attached to such entities? What regulatory frameworks should be in place to regulate their procurement for reproduction or research? And how should such transactions be acknowledged? These are some of the questions that came to my mind when reading Noggle and colleagues' paper¹.

Since the time of Roman law, legal thinking has operated with a fundamental distinction between person and thing. Even today, the entities subject to regulation are either persons or things, and there is no third option⁸. This conceptual lacuna continues to generate regulatory paradoxes in the health and life sciences, because many of the entities subject to regulation — including bodies, body parts, organs and tissues, and sperm and oocytes — cannot be considered either persons or mere things.

How, then, should researchers proceed to procure oocytes? The approach Noggle *et al.* have taken is to pay 16 women for their oocytes and acknowledge their contribution as study participants. I believe this is a step in the right direction for three reasons: first, it transfers the focus from the entities procured to the subjects providing them; second, this refocusing avoids reducing the oocytes to mere things or commodities open for transactions according to the rules of the market; and finally, the word 'participation' paves the way for acknowledging the women's contribution as a piece of work for which they should be duly paid.

The standard argument against paying gamete donors is that the contribution is only material — and therefore marginal — compared with that of the researchers involved. But whether a differential valuation between intellectual input and input of a material or manual kind is justified is questionable. As bioethicist Søren Holm wrote⁹: "In a future situation where there are many groups deriving stem cells, and many donors providing embryos or gametes for the derivation, everyone's contributions will be equally accidental and contingent...". If one group of accidental contributors (the researchers) is entitled to benefit financially from their contribution, why deny payment to another group of accidental contributors (the oocyte providers) for their work?

Another argument against paying oocyte providers is that this would undermine the voluntary nature of the consent process and give an undue incentive to participate in such research¹⁰. This argument also seems to be based on questionable grounds, because the prospect of obtaining future financial benefits from participating in research may also represent a sort of undue inducement for the researchers. Besides, an indication that the women involved in the present study¹ did not necessarily participate for financial gain is that they were all fully employed.

The way Noggle *et al.*¹ have chosen to deal with the oocyte issue does not comply neatly

with existing regulatory guidelines in the field of stem-cell research. For this, in my view, they deserve praise rather than criticism, because their approach helps to draw attention to a possible way out of the regulatory quagmire resulting from reduction of oocyte providers to 'donors' or 'gift givers' deserving merely compensation for their gifts. The authors' approach represents the first step towards acknowledging women as genuine participants — co-producers even — in the generation of new knowledge. ■

Jan Helge Solbakk is in the Centre for Medical Ethics, Faculty of Medicine, University of Oslo, Box 1130, Blindern, 0318 Oslo, Norway. e-mail: j.h.solbakk@medisin.uio.no

1. Noggle, S. *et al.* *Nature* **478**, 70–75 (2011).
2. Rideout, W. M. *et al.* *Cell* **109**, 17–27 (2002).
3. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
4. Pera, M. F. *Nature* **471**, 46–47 (2011).
5. Brambrink, T., Hochedlinger, K., Bell, G. & Jaenisch, R. *Proc. Natl Acad. Sci. USA* **103**, 933–938 (2006).
6. Kim, K. *et al.* *Nature* **467**, 285–290 (2010).
7. Byrne, J. A. *et al.* *Nature* **450**, 497–502 (2007).
8. Lobato de Faria, P. in *The Ethics of Research Biobanking* (eds Solbakk, J. H., Holm, S. & Hofmann, B.) 263–276 (Springer, 2009).
9. Holm, S. *J. Bioeth. Inquiry* **3**, 55–68 (2006).
10. Hyun, I. *Nature* **442**, 629–630 (2006).

G.Q.D. declares competing financial interests. See go.nature.com/hx1akx for details.

EXTRASOLAR PLANETS

Homing in on another Earth

The identification of the closest analogue of Earth so far, orbiting another star, suggests that small planets are common, and that the discovery of a candidate habitable planet in an alien star system could be just around the corner.

JACOB BEAN

In the hunt for planets around stars other than the Sun, astronomers' primary objective is to find a planet that is teeming with life. A milestone on the path to this goal is the discovery of Earth-sized planets orbiting their parent stars in the 'habitable zone' — the range of distances from the star at which the temperature would be just right for liquid water to be present on a planet's surface. But which stars harbour such planets, how common are they, and what are their basic characteristics? Although astronomers can't yet answer these questions, a paper to be published in *Astronomy & Astrophysics* by Pepe *et al.*¹ presents the discovery of several planets that

marks a significant step towards changing this impasse*.

Pepe and colleagues detected five small planets orbiting parent stars that are slightly smaller and cooler than the Sun. One of the planets is only 3.6 times the mass of Earth and is in an orbit that teases the inner edge of its host star's habitable zone. This is the closest that astronomers have yet come to finding another Earth. Furthermore, the relative ease with which these and other previously reported small planets have been found by the same group implies that the frequency of such planets around Sun-like stars is on the order of tens of per cent.

The authors made their new discoveries¹

*This article was published online on 28 September 2011.

are essential for embryonic cell division or expression of genes from the somatic genome. Frustratingly, they could not overcome this cleavage arrest unless they left the oocyte genome in place; the cells they derived from the resulting blastocysts were therefore triploid somatic–oocyte pluripotent stem cells. Nonetheless, the authors' sophisticated analysis revealed that the transplanted genome was fully reprogrammed, with no signs of epigenetic memory. Thus, although falling short of its ultimate goal, the paper¹ stands as a stepping stone towards success, and raises the provocative question of how human SCNT-ES cells might perform relative to iPS cells.

George Q. Daley is in the Stem Cell Transplantation Program, Division of Pediatric Hematology/Oncology, Howard Hughes Medical Institute, Children's Hospital Boston, Boston, Massachusetts 02115, USA. e-mail: george.daley@childrens.harvard.edu

Persons versus things

JAN HELGE SOLBAKK

What are oocytes? What is their nature? What conceptual labels should be attached to such entities? What regulatory frameworks should be in place to regulate their procurement for reproduction or research? And how should such transactions be acknowledged? These are some of the questions that came to my mind when reading Noggle and colleagues' paper¹.

Since the time of Roman law, legal thinking has operated with a fundamental distinction between person and thing. Even today, the entities subject to regulation are either persons or things, and there is no third option⁸. This conceptual lacuna continues to generate regulatory paradoxes in the health and life sciences, because many of the entities subject to regulation — including bodies, body parts, organs and tissues, and sperm and oocytes — cannot be considered either persons or mere things.

How, then, should researchers proceed to procure oocytes? The approach Noggle *et al.* have taken is to pay 16 women for their oocytes and acknowledge their contribution as study participants. I believe this is a step in the right direction for three reasons: first, it transfers the focus from the entities procured to the subjects providing them; second, this refocusing avoids reducing the oocytes to mere things or commodities open for transactions according to the rules of the market; and finally, the word 'participation' paves the way for acknowledging the women's contribution as a piece of work for which they should be duly paid.

The standard argument against paying gamete donors is that the contribution is only material — and therefore marginal — compared with that of the researchers involved. But whether a differential valuation between intellectual input and input of a material or manual kind is justified is questionable. As bioethicist Søren Holm wrote⁹: "In a future situation where there are many groups deriving stem cells, and many donors providing embryos or gametes for the derivation, everyone's contributions will be equally accidental and contingent...". If one group of accidental contributors (the researchers) is entitled to benefit financially from their contribution, why deny payment to another group of accidental contributors (the oocyte providers) for their work?

Another argument against paying oocyte providers is that this would undermine the voluntary nature of the consent process and give an undue incentive to participate in such research¹⁰. This argument also seems to be based on questionable grounds, because the prospect of obtaining future financial benefits from participating in research may also represent a sort of undue inducement for the researchers. Besides, an indication that the women involved in the present study¹ did not necessarily participate for financial gain is that they were all fully employed.

The way Noggle *et al.*¹ have chosen to deal with the oocyte issue does not comply neatly

with existing regulatory guidelines in the field of stem-cell research. For this, in my view, they deserve praise rather than criticism, because their approach helps to draw attention to a possible way out of the regulatory quagmire resulting from reduction of oocyte providers to 'donors' or 'gift givers' deserving merely compensation for their gifts. The authors' approach represents the first step towards acknowledging women as genuine participants — co-producers even — in the generation of new knowledge. ■

Jan Helge Solbakk is in the Centre for Medical Ethics, Faculty of Medicine, University of Oslo, Box 1130, Blindern, 0318 Oslo, Norway. e-mail: j.h.solbakk@medisin.uio.no

1. Noggle, S. *et al.* *Nature* **478**, 70–75 (2011).
2. Rideout, W. M. *et al.* *Cell* **109**, 17–27 (2002).
3. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
4. Pera, M. F. *Nature* **471**, 46–47 (2011).
5. Brambrink, T., Hochedlinger, K., Bell, G. & Jaenisch, R. *Proc. Natl Acad. Sci. USA* **103**, 933–938 (2006).
6. Kim, K. *et al.* *Nature* **467**, 285–290 (2010).
7. Byrne, J. A. *et al.* *Nature* **450**, 497–502 (2007).
8. Lobato de Faria, P. in *The Ethics of Research Biobanking* (eds Solbakk, J. H., Holm, S. & Hofmann, B.) 263–276 (Springer, 2009).
9. Holm, S. *J. Bioeth. Inquiry* **3**, 55–68 (2006).
10. Hyun, I. *Nature* **442**, 629–630 (2006).

G.Q.D. declares competing financial interests. See go.nature.com/hx1akx for details.

EXTRASOLAR PLANETS

Homing in on another Earth

The identification of the closest analogue of Earth so far, orbiting another star, suggests that small planets are common, and that the discovery of a candidate habitable planet in an alien star system could be just around the corner.

JACOB BEAN

In the hunt for planets around stars other than the Sun, astronomers' primary objective is to find a planet that is teeming with life. A milestone on the path to this goal is the discovery of Earth-sized planets orbiting their parent stars in the 'habitable zone' — the range of distances from the star at which the temperature would be just right for liquid water to be present on a planet's surface. But which stars harbour such planets, how common are they, and what are their basic characteristics? Although astronomers can't yet answer these questions, a paper to be published in *Astronomy & Astrophysics* by Pepe *et al.*¹ presents the discovery of several planets that

marks a significant step towards changing this impasse*.

Pepe and colleagues detected five small planets orbiting parent stars that are slightly smaller and cooler than the Sun. One of the planets is only 3.6 times the mass of Earth and is in an orbit that teases the inner edge of its host star's habitable zone. This is the closest that astronomers have yet come to finding another Earth. Furthermore, the relative ease with which these and other previously reported small planets have been found by the same group implies that the frequency of such planets around Sun-like stars is on the order of tens of per cent.

The authors made their new discoveries¹

*This article was published online on 28 September 2011.

using the radial-velocity technique, which is an indirect method based on measuring the periodic change in speed of a star caused by the gravitational tug of orbiting bodies. Until this year, most planet discoveries were made with this method. But there has been a widely held opinion that the radial-velocity technique will not be able to find candidate habitable Earth-mass planets despite its success at finding Jupiter-mass planets. The reasoning behind this is that the variability of the visible surfaces of stars, as a result of magnetic activity, pulsations and the turbulence of the plasma in their atmospheres, creates noise in radial-velocity measurements, and this noise is larger than the signal induced by an Earth-sized planet in the habitable zone. Also, the technical challenge of building instruments sensitive enough to detect the subtle indications of such planets, even around perfectly 'quiet' stars, was considered too challenging.

Pepe and colleagues' detections¹ were enabled by a combination of exquisite instrumentation, painstaking analysis and an ingenious observational technique. Using an ultra-stable instrument on a dedicated telescope, and with the benefit of ten years' experience in refining their calibration strategy, the authors set out to intensively monitor ten of the quietest known nearby stars. Their monitoring campaign involved making multiple measurements per night to average over the stellar noise cycles that had plagued previous studies. Using this approach, they have broken through the previously inviolate one-metre-per-second radial-velocity barrier and detected planets that have signals less than ten-fold larger than that of an Earth analogue, for which the signal is 9 cm s^{-1} . Pepe *et al.* were also careful to show that the detected signals were most probably attributable to orbiting planets rather than to intrinsic variations in the stars themselves. They did this by demonstrating that the orbital periods determined for the planets were distinct from the host stars' rotation periods, and also that there were no correlations between the detected signals and the diagnostics of stellar activity.

Despite being a clear breakthrough, there are some limitations to Pepe and colleagues' study. One limitation is that the newly detected planets are poorly characterized. We do not know the compositions of the planets because the radial-velocity method yields no information about the densities of the planets it detects. Also, we have little knowledge of how elliptical the planets' orbits are (their orbital eccentricity), because the detected signals are relatively small. Both composition and orbital eccentricity are crucial parameters for assessing a planet's habitability.

Furthermore, although Pepe and colleagues' results¹, and those described in another paper from the same group², hint at a large population of small planets orbiting Sun-like stars, the statistics are weak because these planets are difficult to detect and therefore the samples are

incomplete by an unknown amount. Identifying a large sample of similar planets, and studying them by different methods, would advance the field. However, Pepe *et al.* show that the full potential of the radial-velocity technique has not yet been realized. This work warrants the building of a generation of radial-

One of the planets is only 3.6 times the mass of Earth and is in an orbit that teases the inner edge of its host star's 'habitable' zone.

velocity instruments that are even more stable than those now in use, so that instrumental noise is no longer a barrier to the discovery of Earth-like planets with this method.

NASA's ongoing Kepler mission, which finds planets using the transit technique, announced a haul of more than 1,000 new planet candidates earlier this year, and is on track to identify habitable-zone planets³. Because the transit technique detects planets by measuring periodic decreases in a star's brightness when a planet passes in front of it, the issue of stellar variability is also a limiting factor for this approach, and early results indicate that the stars Kepler is looking at are significantly more variable than expected⁴. This means that

an extension beyond the nominal 3.5-year mission duration might be necessary for Kepler to securely detect Earth-sized planets in the habitable zones of Sun-like stars.

In the long run, astronomers aim to study the atmospheres of the small worlds revealed by radial-velocity and transit surveys to obtain further insight into the planets' habitability and even, perhaps, their state of inhabitation. The investigations that will be enabled by NASA's planned James Webb Space Telescope, which is currently at risk of cancellation, are the cornerstone of astronomers' next plans in this area. Nevertheless, it is clear that the search for other Earths is gathering pace. The exciting results from Kepler, and the remarkable advances in the radial-velocity technique demonstrated by Pepe *et al.*, show that the race is well and truly on. ■

Jacob Bean is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.
e-mail: jbean@cfa.harvard.edu

1. Pepe, F. *et al.* *Astron. Astrophys.* (in the press); preprint at <http://arxiv.org/abs/1108.3447> (2011).
2. Mayor, M. *et al.* preprint at <http://arxiv.org/abs/1109.2497> (2011).
3. Borucki, W. J. *et al.* *Astrophys. J.* **736**, 19 (2011).
4. Gilliland, R. L. *et al.* *Astrophys. J.* (in the press); preprint at <http://arxiv.org/abs/1107.5207> (2011).

MOLECULAR MEDICINE

Defence against oxidative damage

Macular degeneration is a leading cause of blindness in the elderly in the developed world. Hope for prevention and treatment comes from the discovery of a protective mechanism against oxidative damage to the eye. SEE LETTER P.76

FERNANDO CRUZ-GUILLOT & VICTOR L. PEREZ

Dangerous oxygen radicals that are sometimes generated during metabolic processes can damage cellular components. The eye, with its constant exposure to light and its high metabolic rate, is particularly susceptible to such damage, or oxidative stress. If left unchecked, this can be cumulative, leading to age-related macular degeneration. Almost two-thirds of people over the age of 80 have this condition, and between 30 million and 50 million individuals are affected worldwide, with a frequency in industrialized countries similar to that of cancer¹. On page 76 of this issue, Weismann *et al.*² describe how a protein normally associated with an immune pathway also protects against inflammation induced by oxidative stress in age-related

macular degeneration. By combining *in vitro* and *in vivo* data from human patients and animal models, the authors provide a plausible explanation for the cause of this devastating chronic disease.

The macula region of the retina is required for central vision and is heavily populated with photoreceptors. These convert the light entering the eye into electrical and molecular signals that are transmitted to the brain for visual processing. The retina's outer segments are replenished daily, and the resulting debris is cleared away by the retinal pigment epithelial cells. If these cells become dysfunctional, a build-up of debris (drusen) can occur in the vicinity of photoreceptor cells in the macula, which are then more likely to die off, leading to the irreversible loss of vision seen in patients with age-related macular degeneration (AMD). Many environmental and genetic factors have been

using the radial-velocity technique, which is an indirect method based on measuring the periodic change in speed of a star caused by the gravitational tug of orbiting bodies. Until this year, most planet discoveries were made with this method. But there has been a widely held opinion that the radial-velocity technique will not be able to find candidate habitable Earth-mass planets despite its success at finding Jupiter-mass planets. The reasoning behind this is that the variability of the visible surfaces of stars, as a result of magnetic activity, pulsations and the turbulence of the plasma in their atmospheres, creates noise in radial-velocity measurements, and this noise is larger than the signal induced by an Earth-sized planet in the habitable zone. Also, the technical challenge of building instruments sensitive enough to detect the subtle indications of such planets, even around perfectly 'quiet' stars, was considered too challenging.

Pepe and colleagues' detections¹ were enabled by a combination of exquisite instrumentation, painstaking analysis and an ingenious observational technique. Using an ultra-stable instrument on a dedicated telescope, and with the benefit of ten years' experience in refining their calibration strategy, the authors set out to intensively monitor ten of the quietest known nearby stars. Their monitoring campaign involved making multiple measurements per night to average over the stellar noise cycles that had plagued previous studies. Using this approach, they have broken through the previously inviolate one-metre-per-second radial-velocity barrier and detected planets that have signals less than ten-fold larger than that of an Earth analogue, for which the signal is 9 cm s^{-1} . Pepe *et al.* were also careful to show that the detected signals were most probably attributable to orbiting planets rather than to intrinsic variations in the stars themselves. They did this by demonstrating that the orbital periods determined for the planets were distinct from the host stars' rotation periods, and also that there were no correlations between the detected signals and the diagnostics of stellar activity.

Despite being a clear breakthrough, there are some limitations to Pepe and colleagues' study. One limitation is that the newly detected planets are poorly characterized. We do not know the compositions of the planets because the radial-velocity method yields no information about the densities of the planets it detects. Also, we have little knowledge of how elliptical the planets' orbits are (their orbital eccentricity), because the detected signals are relatively small. Both composition and orbital eccentricity are crucial parameters for assessing a planet's habitability.

Furthermore, although Pepe and colleagues' results¹, and those described in another paper from the same group², hint at a large population of small planets orbiting Sun-like stars, the statistics are weak because these planets are difficult to detect and therefore the samples are

incomplete by an unknown amount. Identifying a large sample of similar planets, and studying them by different methods, would advance the field. However, Pepe *et al.* show that the full potential of the radial-velocity technique has not yet been realized. This work warrants the building of a generation of radial-

One of the planets is only 3.6 times the mass of Earth and is in an orbit that teases the inner edge of its host star's 'habitable' zone.

velocity instruments that are even more stable than those now in use, so that instrumental noise is no longer a barrier to the discovery of Earth-like planets with this method.

NASA's ongoing Kepler mission, which finds planets using the transit technique, announced a haul of more than 1,000 new planet candidates earlier this year, and is on track to identify habitable-zone planets³. Because the transit technique detects planets by measuring periodic decreases in a star's brightness when a planet passes in front of it, the issue of stellar variability is also a limiting factor for this approach, and early results indicate that the stars Kepler is looking at are significantly more variable than expected⁴. This means that

an extension beyond the nominal 3.5-year mission duration might be necessary for Kepler to securely detect Earth-sized planets in the habitable zones of Sun-like stars.

In the long run, astronomers aim to study the atmospheres of the small worlds revealed by radial-velocity and transit surveys to obtain further insight into the planets' habitability and even, perhaps, their state of inhabitation. The investigations that will be enabled by NASA's planned James Webb Space Telescope, which is currently at risk of cancellation, are the cornerstone of astronomers' next plans in this area. Nevertheless, it is clear that the search for other Earths is gathering pace. The exciting results from Kepler, and the remarkable advances in the radial-velocity technique demonstrated by Pepe *et al.*, show that the race is well and truly on. ■

Jacob Bean is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.
e-mail: jbean@cfa.harvard.edu

1. Pepe, F. *et al.* *Astron. Astrophys.* (in the press); preprint at <http://arxiv.org/abs/1108.3447> (2011).
2. Mayor, M. *et al.* preprint at <http://arxiv.org/abs/1109.2497> (2011).
3. Borucki, W. J. *et al.* *Astrophys. J.* **736**, 19 (2011).
4. Gilliland, R. L. *et al.* *Astrophys. J.* (in the press); preprint at <http://arxiv.org/abs/1107.5207> (2011).

MOLECULAR MEDICINE

Defence against oxidative damage

Macular degeneration is a leading cause of blindness in the elderly in the developed world. Hope for prevention and treatment comes from the discovery of a protective mechanism against oxidative damage to the eye. SEE LETTER P.76

FERNANDO CRUZ-GUILLOT & VICTOR L. PEREZ

Dangerous oxygen radicals that are sometimes generated during metabolic processes can damage cellular components. The eye, with its constant exposure to light and its high metabolic rate, is particularly susceptible to such damage, or oxidative stress. If left unchecked, this can be cumulative, leading to age-related macular degeneration. Almost two-thirds of people over the age of 80 have this condition, and between 30 million and 50 million individuals are affected worldwide, with a frequency in industrialized countries similar to that of cancer¹. On page 76 of this issue, Weismann *et al.*² describe how a protein normally associated with an immune pathway also protects against inflammation induced by oxidative stress in age-related

macular degeneration. By combining *in vitro* and *in vivo* data from human patients and animal models, the authors provide a plausible explanation for the cause of this devastating chronic disease.

The macula region of the retina is required for central vision and is heavily populated with photoreceptors. These convert the light entering the eye into electrical and molecular signals that are transmitted to the brain for visual processing. The retina's outer segments are replenished daily, and the resulting debris is cleared away by the retinal pigment epithelial cells. If these cells become dysfunctional, a build-up of debris (drusen) can occur in the vicinity of photoreceptor cells in the macula, which are then more likely to die off, leading to the irreversible loss of vision seen in patients with age-related macular degeneration (AMD). Many environmental and genetic factors have been

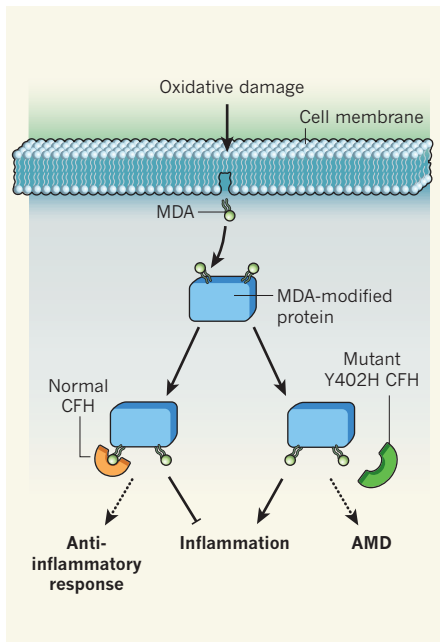


Figure 1 | Mechanisms at work in age-related macular degeneration. Oxidative damage to lipids in the cell membrane generates the reactive decomposition product malondialdehyde (MDA), which forms adducts with cell proteins. Normal complement factor H (CFH) binds MDA with high affinity, blocking inflammatory reactions. Mutant CFH, in which the amino acid histidine is substituted for tyrosine at residue 402, fails to bind MDA, so inflammation cannot be prevented, leading to age-related macular degeneration (AMD) and blindness.

correlated experimentally with AMD progression, including oxidative damage (induced by factors such as smoking and light exposure) and inflammation³.

A collection of proteins known as complement factors form part of the innate immune system, which is the first line of defence against pathogens. These factors interact with one another in a sequence of stimulatory or inhibitory steps in a cascade known as the complement pathway. Complement proteins have been implicated in certain pathological conditions, and have been found in the accumulated drusen of patients with AMD⁴. Variations in the DNA sequence at particular sites, or polymorphisms, in genes encoding complement factors have been associated with the development of AMD, suggesting that inflammation is an important component of the disease.

A polymorphism in complement factor H (CFH) conveys a significant risk of developing AMD^{5–7}. CFH is an inhibitor of the complement pathway and therefore has anti-inflammatory activity. The relevant single-nucleotide polymorphism in the *CFH* gene produces an amino-acid change from tyrosine to histidine at position 402 (the Y402H mutation) in the protein. The functional consequences of this mutation have been elusive until now, but

Weismann *et al.*² convincingly show that it directly affects the ability of CFH to control the inflammation associated with AMD.

The story began with the group's interest in malondialdehyde (MDA) — a common decomposition product of lipid peroxidation by oxygen radicals. It reacts with cellular proteins to form adducts that can act as markers of oxidative stress. The MDA-modified proteins induce inflammatory responses and are recognized by the innate immune system. They are found in many physiological and pathological conditions, including atherosclerosis, AMD^{8,9} and other chronic degenerative diseases.

Weismann *et al.* show that CFH peptides constitute the majority of MDA-binding proteins. A series of cleverly designed experiments clarified the physical and functional features of the CFH–MDA interaction. This turned out to be highly specific, with CFH binding to MDA whatever its carrier protein, but not to other oxidative products. The authors mapped the CFH domains that specified MDA binding, including a short segment known as SCR7, which contains the Y402H mutation. The Y402H CFH variant from AMD patients showed a markedly reduced ability to bind MDA compared with normal CFH (Fig. 1).

Weismann and colleagues further demonstrate that CFH and MDA co-localize in the eyes irrespective of whether these organs are affected by AMD. This suggests that in the healthy eye CFH protects the macula, and in dying (apoptotic) cells it recognizes MDA–protein adducts. Perhaps the researchers' most important result from a therapeutic viewpoint is that CFH can prevent MDA-mediated pro-inflammatory effects in at least two cell types associated with AMD — retinal pigment epithelial cells and macrophages.

Weismann and co-workers' findings² answer the long-standing question about the role of CFH in AMD but they also raise other questions. MDA is ubiquitously generated in a variety of inflammatory settings, but we don't know whether its connection with CFH is relevant outside the eye. The authors found that other members of the CFH family that bind MDA block CFH activity, suggesting that the subtle regulation of complement activity needs to be examined in more detail. It will be interesting to see whether other oxidation-induced modifications associated with AMD, including those induced by carboxyethylpyrrole¹⁰, interact with proteins in a way similar to the MDA–CFH paradigm. Answers to such questions could help in the fight against AMD and other chronic inflammatory diseases. ■

Fernando Cruz-Guilloty and Victor L. Perez are in the Department of Ophthalmology, Bascom Palmer Eye Institute, and the Department of Microbiology and Immunology, University of Miami

Miller School of Medicine, Miami, Florida 33136, USA.
e-mail: vperez4@med.miami.edu

1. Takeda, A. *et al.* *Nature* **460**, 225–230 (2009).
2. Weismann, D. *et al.* *Nature* **478**, 76–81 (2011).
3. Telander, D. G. *Semin. Ophthalmol.* **26**, 192–197 (2011).

4. Crabb, J. W. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 14682–14687 (2002).
5. Klein, R. J. *et al.* *Science* **308**, 385–389 (2005).
6. Haines, J. L. *et al.* *Science* **308**, 419–421 (2005).
7. Edwards, A. O. *et al.* *Science* **308**, 421–424 (2005).
8. Miller, Y. I. *et al.* *Circ. Res.* **108**, 235–248 (2011).
9. Suzuki, M. *et al.* *Mol. Vision* **13**, 772–778 (2007).
10. Hollyfield, J. G. *et al.* *Nature Med.* **14**, 194–198 (2008).

PALAEOANTHROPOLOGY

Malapa and the genus *Homo*

Two remarkably well-preserved skeletons of the hominin species *Australopithecus sediba*, found at Malapa, South Africa, show an intriguing combination of features, and open up a debate about the origins of the genus *Homo*.

FRED SPOOR

Following on from the announcement last year by Berger *et al.*¹ of the remains of a newly discovered hominin species, *Australopithecus sediba*, the same group has now published five reports^{2–6} in *Science* detailing additional fossils and further analyses. Cave deposits at the Malapa site in South Africa yielded two partial skeletons, which Pickering *et al.*⁶ have found to be 1.977 ± 0.002 million years (Myr) old. These skeletons are not only well preserved and remarkably complete, but also show a surprising mix of morphological characters. Given the completeness of the skeletons, the unexpected combination of primitive and derived morphology, and the likelihood that further individuals will be recovered at Malapa, *A. sediba* certainly has the potential to uproot conventional views of human evolution.

Overall, the authors find that *A. sediba* is australopith-like, with a small brain and long arms, and is most similar to its likely ancestor *Australopithecus africanus*, remains of which have been found at several South African sites. However, some aspects of the *A. sediba* skeletons seem to show a closer resemblance to the morphology found in species of the genus *Homo*. These include aspects of the shape of the pelvis³ and ankle joint⁵, as well as the long thumb and short fingers that are characteristic of hands capable of precise manipulation⁴. The authors suggest that these features are phylogenetically shared with *Homo* species, rather than being examples of homoplasy (similar traits that

evolved independently in separate lineages), and conclude that *A. sediba* is a plausible candidate ancestor of *Homo*.

Early species of the genus *Homo* — *H. habilis*, *H. rudolfensis* and *H. erectus* — appear in the fossil record about 1.9 Myr ago (Fig. 1). Of these, *H. erectus* stands out because it gave rise to later *Homo* species (including modern humans), dispersed out of Africa, and became extinct less than half a million years ago. *Homo habilis*, or a species similar to it,

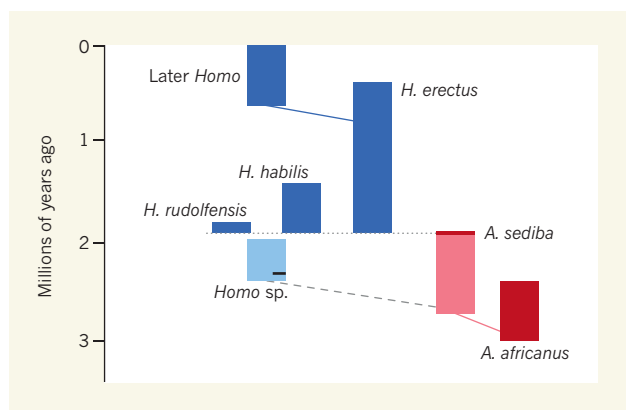


Figure 1 | Temporal distribution of selected hominin species. The bar diagram shows when various hominins (two australopiths, red, and various *Homo* species, blue) appear in the fossil record. The pale blue bar represents fragmentary fossils that are generally thought to come from early *Homo*. Of these, an upper jawbone from Hadar, Ethiopia (black line on the pale blue bar), is well dated at 2.35 million years (Myr) old, and is the most convincingly *Homo*-like. Lines connecting bars indicate hypothetical ancestry between species. The most recent addition to the diagram is *Australopithecus sediba* — two skeletons^{2–6} of the hominin found at Malapa, South Africa, are 1.977 ± 0.002 Myr old. Two scenarios have been proposed^{1,6} in which *A. sediba* is the ancestor of the genus *Homo*. In the first scenario¹, fossils at Malapa come from a late-surviving population of *A. sediba*, whose earlier representatives (pink bar) were ancestral to *Homo* (dashed line). In the second scenario⁶, the *A. sediba* population at Malapa was itself ancestral to early *Homo* (dotted line), which means that fossils pre-dating 2 Myr ago (pale blue) cannot be attributed to *Homo*.

is commonly considered to have been the ancestor of *H. erectus*, but it is difficult to be sure of this because only a small number of fragmentary fossils older than 1.9 Myr have been attributed to *Homo*, and these could not be attributed definitively to a specific species. The fossil most secure in its affinities and provenance is the approximately 2.35-Myr-old upper jawbone from Hadar, Ethiopia⁷, which is more *Homo*-like than that of *A. sediba* and pre-dates the Malapa finds by some 370,000 years. This evidence seems at odds with the idea that *A. sediba* was involved in the first appearance of *Homo*.

In their original publication¹, Berger *et al.* suggested that *A. sediba* could have originated much earlier than the time to which the remains were dated, with the Malapa sample representing a late-surviving population (Fig. 1). However, in their latest report⁶, the authors go further by concluding that even *A. sediba* fossils as late as those preserved at Malapa could have been the ancestor of *Homo*. As a logical corollary, they also contest the *Homo* affinities of any fossil older than 2.0 Myr old. What's more, the authors hint at the possibility that *A. sediba* itself, rather than a species such as *H. habilis* or *H. rudolfensis*, was ancestral to *H. erectus*. It will, however, be difficult to uphold the suggestion that the extensive evolutionary change required could have occurred in the time available (a maximum of 80,000 years) if *A. sediba* at Malapa gave rise to *Homo* species. Moreover, the idea that no fossil older than 2.0 Myr is legitimately attributable to *Homo* is highly debatable — the arguments provided in the paper are insufficiently specific to be conclusive, particularly with respect to the Hadar jawbone.

Another question is whether the authors' morphological analyses^{1–6} do indeed suggest that *A. sediba* has closer evolutionary links with *H. erectus* than do *H. habilis* or *H. rudolfensis*. The answer seems to be no, mainly because the required comparisons either were not made or cannot be made in the absence of fossil evidence. For example, the morphology of the entire post-cranial skeleton of *H. rudolfensis* is unknown, as is that of the pelvis of *H. habilis*, and very few hand and foot bones of *H. erectus* have been recovered, which means that none of these bones can be compared with those of *A. sediba*. Conversely, several brain endocasts — casts of the inside of fossil braincases — of species of early *Homo* are available⁸, but the authors compared² their endocast of *A. sediba* only with those of modern humans and chimpanzees, and with two *A. africanus* fossils.

The rear and base of the *A. sediba*

Miller School of Medicine, Miami, Florida 33136, USA.
e-mail: vperez4@med.miami.edu

1. Takeda, A. *et al.* *Nature* **460**, 225–230 (2009).
2. Weismann, D. *et al.* *Nature* **478**, 76–81 (2011).
3. Telander, D. G. *Semin. Ophthalmol.* **26**, 192–197 (2011).

4. Crabb, J. W. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 14682–14687 (2002).
5. Klein, R. J. *et al.* *Science* **308**, 385–389 (2005).
6. Haines, J. L. *et al.* *Science* **308**, 419–421 (2005).
7. Edwards, A. O. *et al.* *Science* **308**, 421–424 (2005).
8. Miller, Y. I. *et al.* *Circ. Res.* **108**, 235–248 (2011).
9. Suzuki, M. *et al.* *Mol. Vision* **13**, 772–778 (2007).
10. Hollyfield, J. G. *et al.* *Nature Med.* **14**, 194–198 (2008).

PALAEOANTHROPOLOGY

Malapa and the genus *Homo*

Two remarkably well-preserved skeletons of the hominin species *Australopithecus sediba*, found at Malapa, South Africa, show an intriguing combination of features, and open up a debate about the origins of the genus *Homo*.

FRED SPOOR

Following on from the announcement last year by Berger *et al.*¹ of the remains of a newly discovered hominin species, *Australopithecus sediba*, the same group has now published five reports^{2–6} in *Science* detailing additional fossils and further analyses. Cave deposits at the Malapa site in South Africa yielded two partial skeletons, which Pickering *et al.*⁶ have found to be 1.977 ± 0.002 million years (Myr) old. These skeletons are not only well preserved and remarkably complete, but also show a surprising mix of morphological characters. Given the completeness of the skeletons, the unexpected combination of primitive and derived morphology, and the likelihood that further individuals will be recovered at Malapa, *A. sediba* certainly has the potential to uproot conventional views of human evolution.

Overall, the authors find that *A. sediba* is australopith-like, with a small brain and long arms, and is most similar to its likely ancestor *Australopithecus africanus*, remains of which have been found at several South African sites. However, some aspects of the *A. sediba* skeletons seem to show a closer resemblance to the morphology found in species of the genus *Homo*. These include aspects of the shape of the pelvis³ and ankle joint⁵, as well as the long thumb and short fingers that are characteristic of hands capable of precise manipulation⁴. The authors suggest that these features are phylogenetically shared with *Homo* species, rather than being examples of homoplasy (similar traits that

evolved independently in separate lineages), and conclude that *A. sediba* is a plausible candidate ancestor of *Homo*.

Early species of the genus *Homo* — *H. habilis*, *H. rudolfensis* and *H. erectus* — appear in the fossil record about 1.9 Myr ago (Fig. 1). Of these, *H. erectus* stands out because it gave rise to later *Homo* species (including modern humans), dispersed out of Africa, and became extinct less than half a million years ago. *Homo habilis*, or a species similar to it,

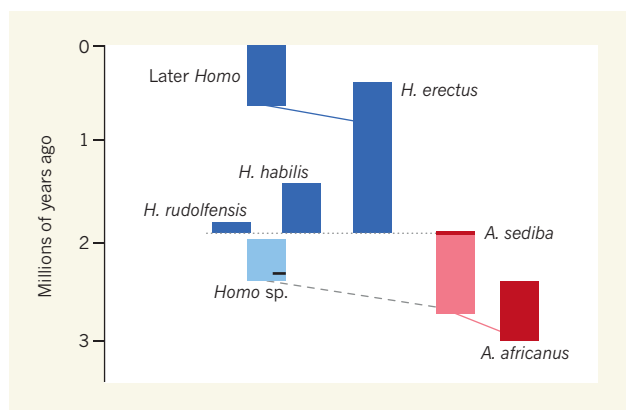


Figure 1 | Temporal distribution of selected hominin species. The bar diagram shows when various hominins (two australopiths, red, and various *Homo* species, blue) appear in the fossil record. The pale blue bar represents fragmentary fossils that are generally thought to come from early *Homo*. Of these, an upper jawbone from Hadar, Ethiopia (black line on the pale blue bar), is well dated at 2.35 million years (Myr) old, and is the most convincingly *Homo*-like. Lines connecting bars indicate hypothetical ancestry between species. The most recent addition to the diagram is *Australopithecus sediba* — two skeletons^{2–6} of the hominin found at Malapa, South Africa, are 1.977 ± 0.002 Myr old. Two scenarios have been proposed^{1,6} in which *A. sediba* is the ancestor of the genus *Homo*. In the first scenario¹, fossils at Malapa come from a late-surviving population of *A. sediba*, whose earlier representatives (pink bar) were ancestral to *Homo* (dashed line). In the second scenario⁶, the *A. sediba* population at Malapa was itself ancestral to early *Homo* (dotted line), which means that fossils pre-dating 2 Myr ago (pale blue) cannot be attributed to *Homo*.

is commonly considered to have been the ancestor of *H. erectus*, but it is difficult to be sure of this because only a small number of fragmentary fossils older than 1.9 Myr have been attributed to *Homo*, and these could not be attributed definitively to a specific species. The fossil most secure in its affinities and provenance is the approximately 2.35-Myr-old upper jawbone from Hadar, Ethiopia⁷, which is more *Homo*-like than that of *A. sediba* and pre-dates the Malapa finds by some 370,000 years. This evidence seems at odds with the idea that *A. sediba* was involved in the first appearance of *Homo*.

In their original publication¹, Berger *et al.* suggested that *A. sediba* could have originated much earlier than the time to which the remains were dated, with the Malapa sample representing a late-surviving population (Fig. 1). However, in their latest report⁶, the authors go further by concluding that even *A. sediba* fossils as late as those preserved at Malapa could have been the ancestor of *Homo*. As a logical corollary, they also contest the *Homo* affinities of any fossil older than 2.0 Myr old. What's more, the authors hint at the possibility that *A. sediba* itself, rather than a species such as *H. habilis* or *H. rudolfensis*, was ancestral to *H. erectus*. It will, however, be difficult to uphold the suggestion that the extensive evolutionary change required could have occurred in the time available (a maximum of 80,000 years) if *A. sediba* at Malapa gave rise to *Homo* species. Moreover, the idea that no fossil older than 2.0 Myr is legitimately attributable to *Homo* is highly debatable — the arguments provided in the paper are insufficiently specific to be conclusive, particularly with respect to the Hadar jawbone.

Another question is whether the authors' morphological analyses^{1–6} do indeed suggest that *A. sediba* has closer evolutionary links with *H. erectus* than do *H. habilis* or *H. rudolfensis*. The answer seems to be no, mainly because the required comparisons either were not made or cannot be made in the absence of fossil evidence. For example, the morphology of the entire post-cranial skeleton of *H. rudolfensis* is unknown, as is that of the pelvis of *H. habilis*, and very few hand and foot bones of *H. erectus* have been recovered, which means that none of these bones can be compared with those of *A. sediba*. Conversely, several brain endocasts — casts of the inside of fossil braincases — of species of early *Homo* are available⁸, but the authors compared² their endocast of *A. sediba* only with those of modern humans and chimpanzees, and with two *A. africanus* fossils.

The rear and base of the *A. sediba*

STATISTICAL PHYSICS

Self-aware particles

The signature of the self-interactions that a colloid in solution undergoes has been observed. The observation has implications for single-particle studies of soft matter and biological systems. SEE LETTER P.85

ULRICH F. KEYSER

cranium are not preserved, which is unfortunate as these areas are highly diagnostic of *H. erectus*. However, Berger *et al.*¹ have argued that the two species share two characteristic features in other parts of the cranium that are not present in *H. habilis* or *H. rudolfensis*. One feature is a slightly swollen area under the eye socket, but neither the definition nor the expression of this character in hominins is well established. The other feature is the amount of constriction of the braincase relative to the breadth of the face: both *A. sediba* and *H. erectus* show less constriction than other early hominin species.

At first sight, this shared characteristic does seem to be convincing evidence supporting a link between *A. sediba* and *H. erectus*. However, the constriction of an individual's braincase changes significantly, late in development; Berger and colleagues¹ studied a juvenile *A. sediba* that would have developed greater constriction had it lived to adulthood. Furthermore, early *H. erectus* shows greater constriction than do geologically later forms, and in this respect is similar to *H. habilis*. When all of these factors are taken into account, the exclusive grouping of *A. sediba* with *H. erectus* no longer seems clear.

Taken together, the published evidence^{1–6} indicates that *A. sediba* is a late australopith that has several intriguing *Homo*-like features. If these features do indeed associate *A. sediba* with the emergence of *Homo*, rather than reflecting homoplasy, then it seems that the scenario in which the Malapa specimens represent a late surviving population¹ is the most plausible explanation for Berger and colleagues' findings.

Many reviews of palaeontological research end with the statement that it would be highly desirable to recover more fossils. In this case, however, the Malapa team has already done that. The interpretation of their findings may be a matter of debate, but they have undoubtedly added a spectacular and thought-provoking sample to the hominin fossil record. This achievement represents a major contribution to the study of human evolution in all its complexity. ■

Fred Spoor is in the Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany, and at University College London, UK.
e-mail: f.spoor@eva.mpg.de

- Berger, L. R. *et al.* *Science* **328**, 195–204 (2010).
- Carlson, K. J. *et al.* *Science* **333**, 1402–1407 (2011).
- Kibii, J. M. *et al.* *Science* **333**, 1407–1411 (2011).
- Kivell, T. L. *et al.* *Science* **333**, 1411–1417 (2011).
- Zipfel, B. *et al.* *Science* **333**, 1417–1420 (2011).
- Pickering, R. *et al.* *Science* **333**, 1421–1423 (2011).
- Kimbel, W. H., Johanson, D. C. & Rak, Y. *Am. J. Phys. Anthropol.* **103**, 235–262 (1997).
- Holloway, R. L., Broadfield, D. C. & Yuan, M. S. *The Human Fossil Record* Vol. 3 (Wiley, 2004).

Have you ever floated a small ball on the surface of a body of water? If you have, you'll probably know that, no matter how careful you are, the ball usually moves because of the waves created by its impact on the water surface. The ball moves around until all of the waves' energy has dissipated and the water becomes still. During this process, the ball interacts with the water and with itself. This striking macroscopic phenomenon should be valid for all situations in which an object moves in an incompressible medium. Any particle completely immersed in a liquid should also strongly self-interact. However, both models of and experiments on particle fluctuations on micrometre-length scales often dismiss this hydrodynamic self-interaction — which can be justified if water motion stops on timescales that are experimentally inaccessible. However, on page 85 of this issue, Franosch *et al.*¹ show that it is possible to detect the hydrodynamic self-interaction of a particle in solution undergoing Brownian motion.

Franosch and colleagues monitored

the Brownian motion of a single colloid (a micrometre-sized sphere) in liquid held in a single-beam optical trap (Fig. 1a). Optical traps can confine a single colloid to a region spanning just a few tens of nanometres in all three dimensions^{2,3}. They are easily created by tightly focusing a laser beam, which essentially transforms the colloid in the trap into a harmonic oscillator. The colloid's Brownian fluctuations, which are caused by the impact of the surrounding solvent molecules, can then be monitored (Fig. 1b). In the classical 'overdamped' regime, one would expect the fluctuations to be uncorrelated with one another and hence to display a white power spectral density — that is, the magnitude (power) of the fluctuations as a function of frequency is constant at small frequencies (Fig. 1c). However, the power spectral density should change when the colloid's hydrodynamic self-interactions become relevant and, as a result, the fluctuations are correlated in time.

To observe this effect, Franosch and colleagues¹ tuned two timescales that govern colloidal motion in the optical trap. The first is the relaxation time of the harmonic

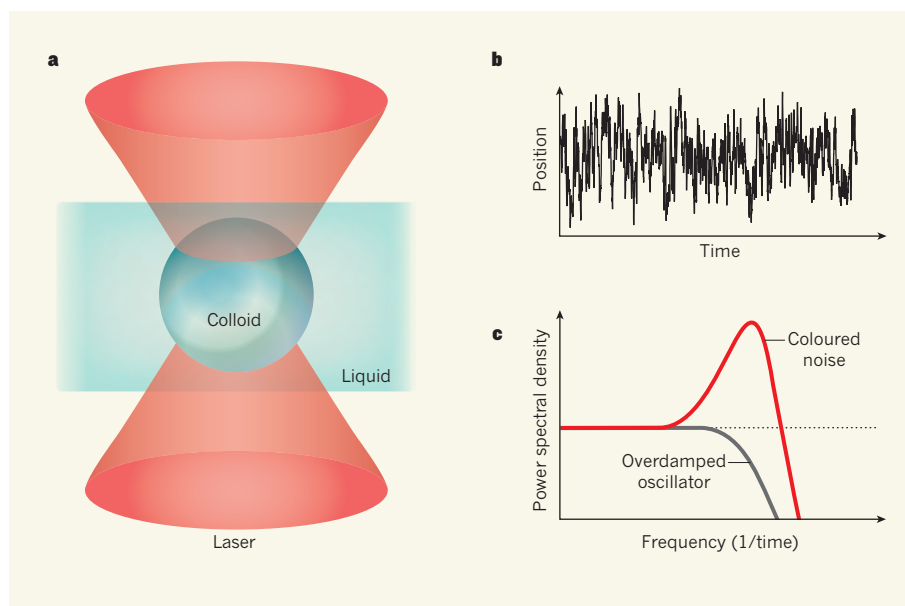


Figure 1 | Coloured noise. **a**, Franosch *et al.*¹ immersed a single colloid in a liquid and trapped it in the focal spot of a laser. **b**, They then monitored the colloid's Brownian fluctuations in position as a function of time. **c**, The resulting dimensionless power spectral density — the magnitude of the fluctuations as a function of frequency (inverse of time) — displayed an increase (coloured noise; red) due to hydrodynamic self-interactions; the classical 'overdamped' regime of oscillations (grey) lacks this increase.

STATISTICAL PHYSICS

Self-aware particles

The signature of the self-interactions that a colloid in solution undergoes has been observed. The observation has implications for single-particle studies of soft matter and biological systems. SEE LETTER P.85

ULRICH F. KEYSER

cranium are not preserved, which is unfortunate as these areas are highly diagnostic of *H. erectus*. However, Berger *et al.*¹ have argued that the two species share two characteristic features in other parts of the cranium that are not present in *H. habilis* or *H. rudolfensis*. One feature is a slightly swollen area under the eye socket, but neither the definition nor the expression of this character in hominins is well established. The other feature is the amount of constriction of the braincase relative to the breadth of the face: both *A. sediba* and *H. erectus* show less constriction than other early hominin species.

At first sight, this shared characteristic does seem to be convincing evidence supporting a link between *A. sediba* and *H. erectus*. However, the constriction of an individual's braincase changes significantly, late in development; Berger and colleagues¹ studied a juvenile *A. sediba* that would have developed greater constriction had it lived to adulthood. Furthermore, early *H. erectus* shows greater constriction than do geologically later forms, and in this respect is similar to *H. habilis*. When all of these factors are taken into account, the exclusive grouping of *A. sediba* with *H. erectus* no longer seems clear.

Taken together, the published evidence^{1–6} indicates that *A. sediba* is a late australopith that has several intriguing *Homo*-like features. If these features do indeed associate *A. sediba* with the emergence of *Homo*, rather than reflecting homoplasy, then it seems that the scenario in which the Malapa specimens represent a late surviving population¹ is the most plausible explanation for Berger and colleagues' findings.

Many reviews of palaeontological research end with the statement that it would be highly desirable to recover more fossils. In this case, however, the Malapa team has already done that. The interpretation of their findings may be a matter of debate, but they have undoubtedly added a spectacular and thought-provoking sample to the hominin fossil record. This achievement represents a major contribution to the study of human evolution in all its complexity. ■

Fred Spoor is in the Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany, and at University College London, UK.
e-mail: f.spoor@eva.mpg.de

- Berger, L. R. *et al.* *Science* **328**, 195–204 (2010).
- Carlson, K. J. *et al.* *Science* **333**, 1402–1407 (2011).
- Kibii, J. M. *et al.* *Science* **333**, 1407–1411 (2011).
- Kivell, T. L. *et al.* *Science* **333**, 1411–1417 (2011).
- Zipfel, B. *et al.* *Science* **333**, 1417–1420 (2011).
- Pickering, R. *et al.* *Science* **333**, 1421–1423 (2011).
- Kimbel, W. H., Johanson, D. C. & Rak, Y. *Am. J. Phys. Anthropol.* **103**, 235–262 (1997).
- Holloway, R. L., Broadfield, D. C. & Yuan, M. S. *The Human Fossil Record* Vol. 3 (Wiley, 2004).

Have you ever floated a small ball on the surface of a body of water? If you have, you'll probably know that, no matter how careful you are, the ball usually moves because of the waves created by its impact on the water surface. The ball moves around until all of the waves' energy has dissipated and the water becomes still. During this process, the ball interacts with the water and with itself. This striking macroscopic phenomenon should be valid for all situations in which an object moves in an incompressible medium. Any particle completely immersed in a liquid should also strongly self-interact. However, both models of and experiments on particle fluctuations on micrometre-length scales often dismiss this hydrodynamic self-interaction — which can be justified if water motion stops on timescales that are experimentally inaccessible. However, on page 85 of this issue, Franosch *et al.*¹ show that it is possible to detect the hydrodynamic self-interaction of a particle in solution undergoing Brownian motion.

Franosch and colleagues monitored

the Brownian motion of a single colloid (a micrometre-sized sphere) in liquid held in a single-beam optical trap (Fig. 1a). Optical traps can confine a single colloid to a region spanning just a few tens of nanometres in all three dimensions^{2,3}. They are easily created by tightly focusing a laser beam, which essentially transforms the colloid in the trap into a harmonic oscillator. The colloid's Brownian fluctuations, which are caused by the impact of the surrounding solvent molecules, can then be monitored (Fig. 1b). In the classical 'overdamped' regime, one would expect the fluctuations to be uncorrelated with one another and hence to display a white power spectral density — that is, the magnitude (power) of the fluctuations as a function of frequency is constant at small frequencies (Fig. 1c). However, the power spectral density should change when the colloid's hydrodynamic self-interactions become relevant and, as a result, the fluctuations are correlated in time.

To observe this effect, Franosch and colleagues¹ tuned two timescales that govern colloidal motion in the optical trap. The first is the relaxation time of the harmonic

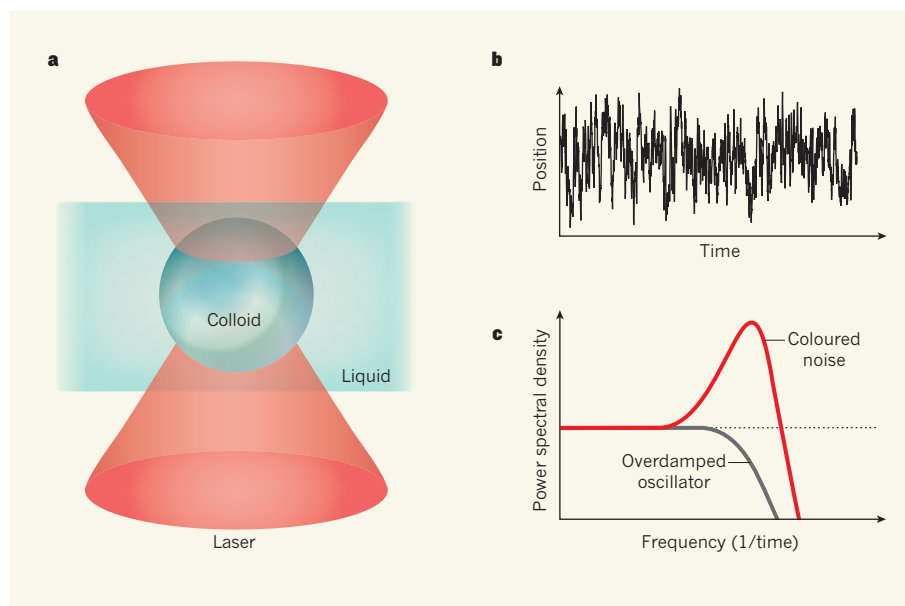


Figure 1 | Coloured noise. **a**, Franosch *et al.*¹ immersed a single colloid in a liquid and trapped it in the focal spot of a laser. **b**, They then monitored the colloid's Brownian fluctuations in position as a function of time. **c**, The resulting dimensionless power spectral density — the magnitude of the fluctuations as a function of frequency (inverse of time) — displayed an increase (coloured noise; red) due to hydrodynamic self-interactions; the classical 'overdamped' regime of oscillations (grey) lacks this increase.

oscillator (the time it takes the colloid to reach its equilibrium position), which is inversely proportional to the trap stiffness, or strength. Because trap stiffness increases with laser power, the authors could easily tune the relaxation time. The second tuned timescale is the diffusion time — the time taken by the fluid to diffuse over the diameter of the particle. This time can be controlled by varying the particle's diameter and the fluid's viscosity. To optimize their measurement conditions, Franosch *et al.* used colloids made of melamine resin with diameters of around 3 micrometres and immersed them in acetone.

By using high time resolution in particle detection and high laser power, the authors managed to decrease the relaxation time to just six times that of the faster fluid diffusion time. This allowed them to observe correlated fluctuations due to the colloid's hydrodynamic self-interaction. A direct consequence of this hitherto undetected correlated behaviour is an increase (resonance) in the power spectral density at frequencies close to the inverse of the relaxation time. Franosch *et al.* show that the resulting power spectral density is not white but 'coloured', because not all frequencies have the same magnitude (Fig. 1c). Interestingly, as shown by their computer models, the observed increase in the power spectral density depends on a particle's shape and surface properties.

Fransoch and colleagues' study has several ramifications for the interpretation of single-particle studies of soft matter. Hydrodynamic coupling between colloids and macromolecules is well known to play a part in multi-particle systems because of the long-range nature of the interaction. For instance, it can be used to explain the coordinated motion of pairs of flagella in *Chlamydomonas* algae⁴. Efforts are also under way to investigate the coupling of several colloids with a view to creating an artificial and autonomous swimmer⁵. In addition to their relevance on micrometre-length scales, hydrodynamic interactions are crucial for understanding phenomena such as the transport of DNA through nanopores under the influence of an electric field⁶.

The results¹ discussed here are remarkable. For the first time, the self-interaction between a single colloid and its surrounding medium is conclusively demonstrated. From these experiments, it seems clear that a single particle in solution is aware of its own presence. Such awareness fundamentally changes the particle's thermal fluctuations, which are mediated by hydrodynamic self-interactions. These findings highlight the fact that hydrodynamic phenomena, often dismissed at the micrometre and nanometre scales in viscous solutions, have to be taken into account and can even be exploited for new sensing applications. One obvious, truly label-free sensing application could be to distinguish shape, changes in diameter and surface

morphology of small particles — or even living cells — in optical traps by monitoring the shape of the power spectral density. ■

Ulrich F. Keyser is in the Biological and Soft Systems Sector, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.
e-mail: ufk20@cam.ac.uk

MUSCULAR DYSTROPHY

A hidden ancestral legacy trumped

A previously unsuspected genetic mechanism underlies a type of muscular dystrophy common in Japan. A therapeutic approach based on this finding and tested in mice has come up with encouraging results. [SEE LETTER P127](#)

MASAYUKI NAKAMORI & CHARLES THORNTON

The symptoms of a genetic disorder known as Fukuyama-type congenital muscular dystrophy start in infancy and lead to severe disability and premature death in childhood or adolescence¹. It is one of the most common recessive diseases in Japan and results from disruption of the gene encoding a protein called fukutin. On page 127 of this issue, Taniguchi-Ikeda *et al.*² identify and correct the

molecular steps that result in the expression of faulty fukutin by the damaged gene.

About 100 generations ago, a segment of DNA in a Japanese ancestor was copied from one genomic site and reinserted into another, landing in the *fukutin* gene³. Genetic events of this kind, in which DNA segments called retrotransposons are copied and pasted in many places in the same genome, are not rare, occurring with an estimated frequency of roughly one per 20 births⁴. The immediate

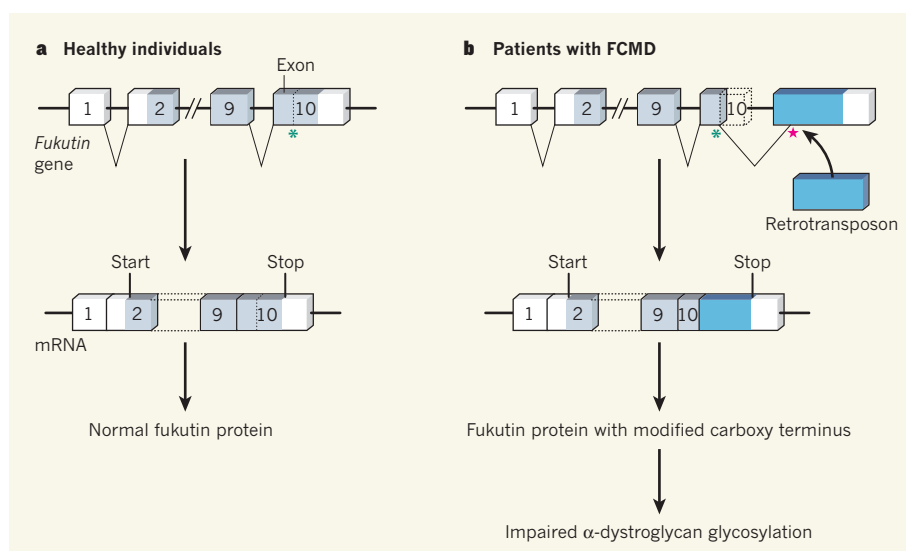


Figure 1 | Abnormal splicing of the *fukutin* gene. **a**, The normal *fukutin* gene contains an inaccessible splice donor site within its exon 10 (green asterisk). The resulting messenger RNA encodes the normal fukutin protein. 'Start' and 'Stop' indicate the beginning and the end of the mRNA sequence encoding the protein. **b**, In patients with Fukuyama-type congenital muscular dystrophy (FCMD), ancestral insertion of a retrotransposon within the final exon activates this inaccessible splice donor site and creates a new splice acceptor site (red star) in the retrotransposon sequence. Such 'exon trapping' results in incorrect splicing of the mRNA and — following translation — modification of the carboxy terminus of the fukutin protein and impaired glycosylation of α-dystroglycan. Taniguchi-Ikeda *et al.*² correct this by using splice-blocking antisense oligonucleotides (not shown) to restore expression of normal fukutin.

oscillator (the time it takes the colloid to reach its equilibrium position), which is inversely proportional to the trap stiffness, or strength. Because trap stiffness increases with laser power, the authors could easily tune the relaxation time. The second tuned timescale is the diffusion time — the time taken by the fluid to diffuse over the diameter of the particle. This time can be controlled by varying the particle's diameter and the fluid's viscosity. To optimize their measurement conditions, Franosch *et al.* used colloids made of melamine resin with diameters of around 3 micrometres and immersed them in acetone.

By using high time resolution in particle detection and high laser power, the authors managed to decrease the relaxation time to just six times that of the faster fluid diffusion time. This allowed them to observe correlated fluctuations due to the colloid's hydrodynamic self-interaction. A direct consequence of this hitherto undetected correlated behaviour is an increase (resonance) in the power spectral density at frequencies close to the inverse of the relaxation time. Franosch *et al.* show that the resulting power spectral density is not white but 'coloured', because not all frequencies have the same magnitude (Fig. 1c). Interestingly, as shown by their computer models, the observed increase in the power spectral density depends on a particle's shape and surface properties.

Fransoch and colleagues' study has several ramifications for the interpretation of single-particle studies of soft matter. Hydrodynamic coupling between colloids and macromolecules is well known to play a part in multi-particle systems because of the long-range nature of the interaction. For instance, it can be used to explain the coordinated motion of pairs of flagella in *Chlamydomonas* algae⁴. Efforts are also under way to investigate the coupling of several colloids with a view to creating an artificial and autonomous swimmer⁵. In addition to their relevance on micrometre-length scales, hydrodynamic interactions are crucial for understanding phenomena such as the transport of DNA through nanopores under the influence of an electric field⁶.

The results¹ discussed here are remarkable. For the first time, the self-interaction between a single colloid and its surrounding medium is conclusively demonstrated. From these experiments, it seems clear that a single particle in solution is aware of its own presence. Such awareness fundamentally changes the particle's thermal fluctuations, which are mediated by hydrodynamic self-interactions. These findings highlight the fact that hydrodynamic phenomena, often dismissed at the micrometre and nanometre scales in viscous solutions, have to be taken into account and can even be exploited for new sensing applications. One obvious, truly label-free sensing application could be to distinguish shape, changes in diameter and surface

morphology of small particles — or even living cells — in optical traps by monitoring the shape of the power spectral density. ■

Ulrich F. Keyser is in the Biological and Soft Systems Sector, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK.
e-mail: ufk20@cam.ac.uk

MUSCULAR DYSTROPHY

A hidden ancestral legacy trumped

A previously unsuspected genetic mechanism underlies a type of muscular dystrophy common in Japan. A therapeutic approach based on this finding and tested in mice has come up with encouraging results. [SEE LETTER P127](#)

MASAYUKI NAKAMORI & CHARLES THORNTON

The symptoms of a genetic disorder known as Fukuyama-type congenital muscular dystrophy start in infancy and lead to severe disability and premature death in childhood or adolescence¹. It is one of the most common recessive diseases in Japan and results from disruption of the gene encoding a protein called fukutin. On page 127 of this issue, Taniguchi-Ikeda *et al.*² identify and correct the

molecular steps that result in the expression of faulty fukutin by the damaged gene.

About 100 generations ago, a segment of DNA in a Japanese ancestor was copied from one genomic site and reinserted into another, landing in the *fukutin* gene³. Genetic events of this kind, in which DNA segments called retrotransposons are copied and pasted in many places in the same genome, are not rare, occurring with an estimated frequency of roughly one per 20 births⁴. The immediate

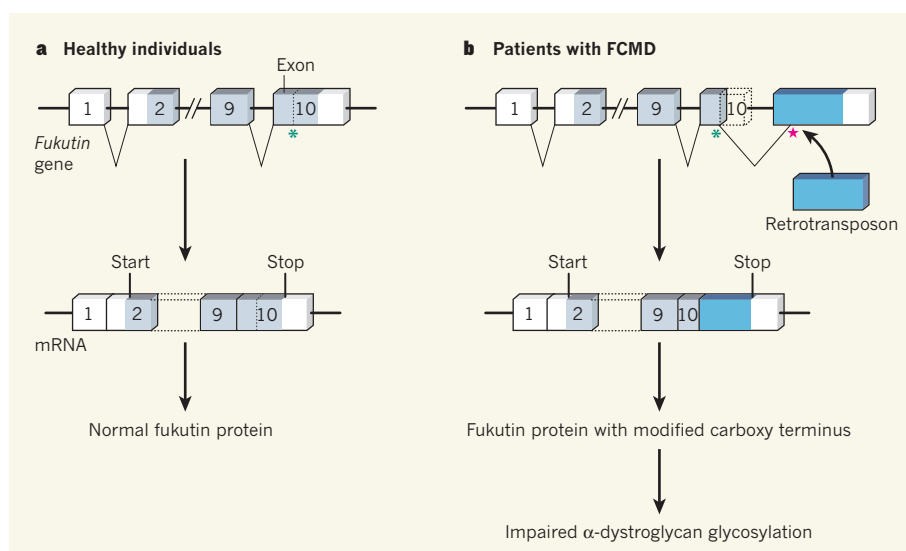


Figure 1 | Abnormal splicing of the *fukutin* gene. **a**, The normal *fukutin* gene contains an inaccessible splice donor site within its exon 10 (green asterisk). The resulting messenger RNA encodes the normal fukutin protein. 'Start' and 'Stop' indicate the beginning and the end of the mRNA sequence encoding the protein. **b**, In patients with Fukuyama-type congenital muscular dystrophy (FCMD), ancestral insertion of a retrotransposon within the final exon activates this inaccessible splice donor site and creates a new splice acceptor site (red star) in the retrotransposon sequence. Such 'exon trapping' results in incorrect splicing of the mRNA and — following translation — modification of the carboxy terminus of the fukutin protein and impaired glycosylation of α-dystroglycan. Taniguchi-Ikeda *et al.*² correct this by using splice-blocking antisense oligonucleotides (not shown) to restore expression of normal fukutin.

descendants of this particular event probably suffered no harm, because disrupting one copy of *fukutin* has no adverse effects.

The consequences for modern Japan, however, are quite different. Around 1 in 90 Japanese individuals now carry this same retrotransposon insertion, being descended from the unknown ancestor. The children of two such individuals are at risk of inheriting two disrupted copies, which results in loss of fukutin activity. Although the exact function of fukutin is unknown, it is clearly involved in the attachment of carbohydrate molecules to the α -dystroglycan protein^{5,6}. This protein is anchored to the cell surface and, when properly modified by carbohydrates through a process called glycosylation, it forms a crucial link between the intracellular cytoskeleton and the extracellular matrix. In the absence of fukutin, glycosylation is incomplete, and the link is broken. This causes abnormal neuronal migration during development, mental retardation and progressive degeneration of muscle cells.

Because the consequences of Fukuyama-type congenital muscular dystrophy (FCMD) are so devastating, the hunt for the disease gene was naturally undertaken with hopes that identification of the mutation would point the way to developing a treatment. But short of using gene therapy to restore a normal copy of the *fukutin* gene, its discovery⁷ had no immediate therapeutic implications. The gene's discoverers continued to pursue the problem, however, and 13 years later they have found a potential opening.

First, they correct a misconception about how the retrotransposon affected *fukutin* expression. Initial studies had indicated that the insertion caused a near-complete absence of *fukutin* messenger RNA, which fitted observations that retrotransposons can silence gene expression. Taniguchi-Ikeda *et al.*² re-examine the problem and find that, although part of the transcript is missing, the overall amount of *fukutin* mRNA is not appreciably reduced. However, they notice that splicing of the *fukutin* transcript, a process in which different parts of the primary transcript are joined to create the mature mRNA, is dramatically affected.

The authors found that the effects on fukutin splicing and function in mice were very similar when they artificially inserted the same retrotransposon at the identical location in the mouse *fukutin* gene. They showed that a splice 'donor' site in the final exon (protein-coding region) that had previously been inaccessible was activated and became joined to a newly created splice 'acceptor' site in the retrotransposon sequence, a process known as exon trapping (Fig. 1). Retrotransposon insertions are known to cause exon trapping⁸, but this is the first example to show a clear association with disease. Because of this splicing alteration, the carboxy terminus of fukutin is eliminated and

replaced instead with amino acids encoded by the retrotransposon sequence. Exactly how this error compromises the glycosylation of α -dystroglycan is unclear, but it may be that the mutant fukutin protein is routed to the wrong cellular compartment².

To test whether normal fukutin expression could be restored by correcting the abnormal splicing, the researchers designed 'antisense' oligonucleotides to suppress exon trapping. These molecules are short DNA-like fragments that bind, according to the rules of nucleic-acid hybridization, to the *fukutin* transcript before it is spliced, thereby favourably altering the outcome of the splicing process. This approach had been used previously to suppress or shift splicing sites in other disease states, including other forms of muscular dystrophy^{9,10}. In cells derived from patients with FCMD, the antisense oligonucleotides had the intended effect of blocking the deleterious splicing event. As predicted, this led to re-expression of normal fukutin protein and re-establishment of the link between α -dystroglycan and extracellular-matrix proteins. Injecting these oligonucleotides into mice carrying the retrotransposon insertion partially restored normal fukutin protein in muscle tissue, again with improved α -dystroglycan glycosylation.

Could this strategy be adopted to treat children with FCMD? Possibly. But rescuing the associated brain malformation would

presumably require treatment *in utero*, a difficult undertaking. The major challenge to the use of antisense oligonucleotides for splice blocking is distributing them into cells in sufficient quantity to influence splicing processes. Although progress has been made in addressing this problem¹⁰, a general solution, applicable for brain and muscle tissue, is not yet available. It is also possible that a similar approach could be applied to other genetic disorders associated with retrotransposon insertion and exon trapping. ■

Masayuki Nakamori and Charles Thornton
are in the Department of Neurology, University of Rochester Medical Center, Rochester, New York 14642, USA.

e-mail: charles_thornton@urmc.rochester.edu

1. Fukuyama, Y., Osawa, M. & Suzuki, H. *Brain Dev.* **3**, 1–29 (1981).
2. Taniguchi-Ikeda, M. *et al. Nature* **478**, 127–131 (2011).
3. Toda, T. & Kobayashi, K. *J. Mol. Med.* **77**, 816–823 (1999).
4. Xing, J. *et al. Genome Res.* **19**, 1516–1526 (2009).
5. Hayashi, Y. K. *et al. Neurology* **57**, 115–121 (2001).
6. Michele, D. E. *et al. Nature* **418**, 417–422 (2002).
7. Kobayashi, K. *et al. Nature* **394**, 388–392 (1998).
8. Hancks, D. C., Ewing, A. D., Chen, J. E., Tokunaga, K. & Kazazian, H. H. Jr *Genome Res.* **19**, 1983–1991 (2009).
9. Dominski, Z. & Kole, R. *Proc. Natl Acad. Sci. USA* **90**, 8673–8677 (1993).
10. Muntoni, F. & Wood, M. J. A. *Nature Rev. Drug Discov.* **10**, 621–637 (2011).

QUANTUM MECHANICS

The gentle cooling touch of light

Laser light has been used to cool a nanomechanical resonator to its lowest energy state. The result opens the door to testing the principles of quantum mechanics and to applications in quantum information processing. SEE LETTER P.89

FLORIAN MARQUARDT

When you face direct sunlight, besides the brightness and heat that you experience, there is a rather subtle effect. The light produces a force pushing at you — admittedly a tiny one, corresponding to the weight of a few grains of sand. In the past few years, however, researchers have learned how to harness these light forces in the nanoworld and to use them to manipulate the mechanical vibrations of small objects, with remarkable results. On page 89 of this issue, Painter and colleagues (Chan *et al.*¹) describe how they have exploited laser light to dampen the motion of a nanomechanical resonator. On entering the quantum regime, the vibrational energy of the resonator is no longer

continuous. Instead, it is in the form of discrete quanta called phonons. The authors' experiment is the first successful attempt of this type to squeeze essentially all the phonons out of the resonator, leaving the system's vibrations in the lowest possible energy state allowed by quantum mechanics — the ground state. Their results finally pave the way for using light to realize many quantum-physical phenomena in such structures.

The force exerted by light, called the radiation pressure force, was first demonstrated a little more than 100 years ago. Radiation forces have been remarkably successful in manipulating the motion of atoms (for example, in laser-cooling them or trapping them within optical lattices produced by the interference of laser beams). They have also been

descendants of this particular event probably suffered no harm, because disrupting one copy of *fukutin* has no adverse effects.

The consequences for modern Japan, however, are quite different. Around 1 in 90 Japanese individuals now carry this same retrotransposon insertion, being descended from the unknown ancestor. The children of two such individuals are at risk of inheriting two disrupted copies, which results in loss of fukutin activity. Although the exact function of fukutin is unknown, it is clearly involved in the attachment of carbohydrate molecules to the α -dystroglycan protein^{5,6}. This protein is anchored to the cell surface and, when properly modified by carbohydrates through a process called glycosylation, it forms a crucial link between the intracellular cytoskeleton and the extracellular matrix. In the absence of fukutin, glycosylation is incomplete, and the link is broken. This causes abnormal neuronal migration during development, mental retardation and progressive degeneration of muscle cells.

Because the consequences of Fukuyama-type congenital muscular dystrophy (FCMD) are so devastating, the hunt for the disease gene was naturally undertaken with hopes that identification of the mutation would point the way to developing a treatment. But short of using gene therapy to restore a normal copy of the *fukutin* gene, its discovery⁷ had no immediate therapeutic implications. The gene's discoverers continued to pursue the problem, however, and 13 years later they have found a potential opening.

First, they correct a misconception about how the retrotransposon affected *fukutin* expression. Initial studies had indicated that the insertion caused a near-complete absence of *fukutin* messenger RNA, which fitted observations that retrotransposons can silence gene expression. Taniguchi-Ikeda *et al.*² re-examine the problem and find that, although part of the transcript is missing, the overall amount of *fukutin* mRNA is not appreciably reduced. However, they notice that splicing of the *fukutin* transcript, a process in which different parts of the primary transcript are joined to create the mature mRNA, is dramatically affected.

The authors found that the effects on fukutin splicing and function in mice were very similar when they artificially inserted the same retrotransposon at the identical location in the mouse *fukutin* gene. They showed that a splice 'donor' site in the final exon (protein-coding region) that had previously been inaccessible was activated and became joined to a newly created splice 'acceptor' site in the retrotransposon sequence, a process known as exon trapping (Fig. 1). Retrotransposon insertions are known to cause exon trapping⁸, but this is the first example to show a clear association with disease. Because of this splicing alteration, the carboxy terminus of fukutin is eliminated and

replaced instead with amino acids encoded by the retrotransposon sequence. Exactly how this error compromises the glycosylation of α -dystroglycan is unclear, but it may be that the mutant fukutin protein is routed to the wrong cellular compartment².

To test whether normal fukutin expression could be restored by correcting the abnormal splicing, the researchers designed 'antisense' oligonucleotides to suppress exon trapping. These molecules are short DNA-like fragments that bind, according to the rules of nucleic-acid hybridization, to the *fukutin* transcript before it is spliced, thereby favourably altering the outcome of the splicing process. This approach had been used previously to suppress or shift splicing sites in other disease states, including other forms of muscular dystrophy^{9,10}. In cells derived from patients with FCMD, the antisense oligonucleotides had the intended effect of blocking the deleterious splicing event. As predicted, this led to re-expression of normal fukutin protein and re-establishment of the link between α -dystroglycan and extracellular-matrix proteins. Injecting these oligonucleotides into mice carrying the retrotransposon insertion partially restored normal fukutin protein in muscle tissue, again with improved α -dystroglycan glycosylation.

Could this strategy be adopted to treat children with FCMD? Possibly. But rescuing the associated brain malformation would

presumably require treatment *in utero*, a difficult undertaking. The major challenge to the use of antisense oligonucleotides for splice blocking is distributing them into cells in sufficient quantity to influence splicing processes. Although progress has been made in addressing this problem¹⁰, a general solution, applicable for brain and muscle tissue, is not yet available. It is also possible that a similar approach could be applied to other genetic disorders associated with retrotransposon insertion and exon trapping. ■

Masayuki Nakamori and Charles Thornton
are in the Department of Neurology, University of Rochester Medical Center, Rochester, New York 14642, USA.

e-mail: charles_thornton@urmc.rochester.edu

1. Fukuyama, Y., Osawa, M. & Suzuki, H. *Brain Dev.* **3**, 1–29 (1981).
2. Taniguchi-Ikeda, M. *et al. Nature* **478**, 127–131 (2011).
3. Toda, T. & Kobayashi, K. *J. Mol. Med.* **77**, 816–823 (1999).
4. Xing, J. *et al. Genome Res.* **19**, 1516–1526 (2009).
5. Hayashi, Y. K. *et al. Neurology* **57**, 115–121 (2001).
6. Michele, D. E. *et al. Nature* **418**, 417–422 (2002).
7. Kobayashi, K. *et al. Nature* **394**, 388–392 (1998).
8. Hancks, D. C., Ewing, A. D., Chen, J. E., Tokunaga, K. & Kazazian, H. H. Jr *Genome Res.* **19**, 1983–1991 (2009).
9. Dominski, Z. & Kole, R. *Proc. Natl Acad. Sci. USA* **90**, 8673–8677 (1993).
10. Muntoni, F. & Wood, M. J. A. *Nature Rev. Drug Discov.* **10**, 621–637 (2011).

QUANTUM MECHANICS

The gentle cooling touch of light

Laser light has been used to cool a nanomechanical resonator to its lowest energy state. The result opens the door to testing the principles of quantum mechanics and to applications in quantum information processing. SEE LETTER P.89

FLORIAN MARQUARDT

When you face direct sunlight, besides the brightness and heat that you experience, there is a rather subtle effect. The light produces a force pushing at you — admittedly a tiny one, corresponding to the weight of a few grains of sand. In the past few years, however, researchers have learned how to harness these light forces in the nanoworld and to use them to manipulate the mechanical vibrations of small objects, with remarkable results. On page 89 of this issue, Painter and colleagues (Chan *et al.*¹) describe how they have exploited laser light to dampen the motion of a nanomechanical resonator. On entering the quantum regime, the vibrational energy of the resonator is no longer

continuous. Instead, it is in the form of discrete quanta called phonons. The authors' experiment is the first successful attempt of this type to squeeze essentially all the phonons out of the resonator, leaving the system's vibrations in the lowest possible energy state allowed by quantum mechanics — the ground state. Their results finally pave the way for using light to realize many quantum-physical phenomena in such structures.

The force exerted by light, called the radiation pressure force, was first demonstrated a little more than 100 years ago. Radiation forces have been remarkably successful in manipulating the motion of atoms (for example, in laser-cooling them or trapping them within optical lattices produced by the interference of laser beams). They have also been

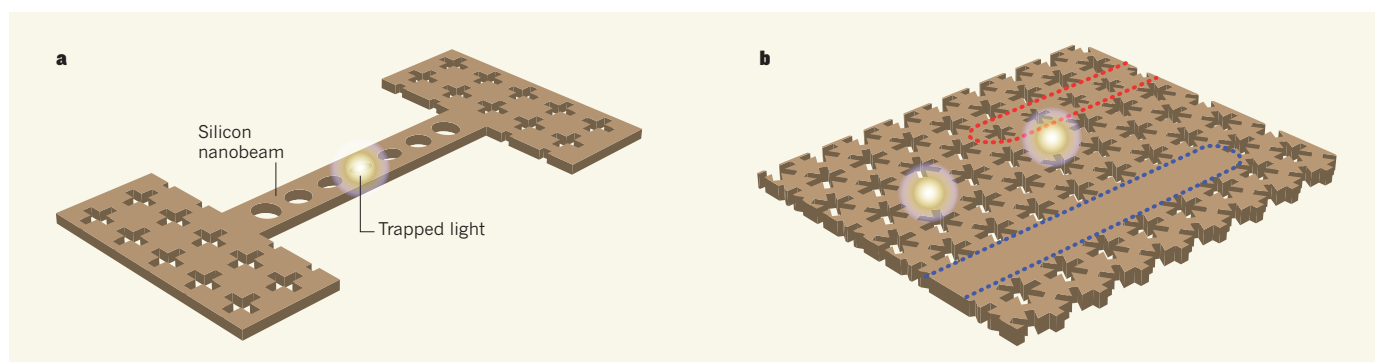


Figure 1 | Coupling light and mechanical motion. **a**, Chan *et al.*¹ patterned a free-standing silicon nanobeam with holes to trap incoming laser light in its central region. This design allowed them to couple the light to the nanobeam's mechanical vibrations (not shown) and bring a particular vibrational standing wave to the quantum-mechanical ground state. **b**, The team already has designs¹⁰ for two-dimensional photonic-crystal

structures — similar to the one shown here — that might form the basis for optomechanical circuits in which light and mechanical motion could be coupled to one another and to optical (blue) and acoustic (red) waveguides. The devices' mechanical and optical functionalities are purely the result of carefully engineering the shapes of the holes cut into the structures.

used to manipulate larger objects such as glass beads, whose motion can be controlled through 'optical tweezers'.

Over the past few years, similar ideas have been applied to control the vibrational motion of nanofabricated structures. A typical set-up involves using a laser to illuminate an optical cavity — an arrangement of two reflective mirrors that allows light to bounce back and forth between them — in which the circulating radiation exerts a force on a mechanical element, such as a vibrating cantilever carrying one of the cavity's mirrors. A large variety of set-ups is being investigated in this rapidly growing field of cavity optomechanics². They involve, for example, not only membranes, microtoroids and nanoscale slabs termed nanobeams, but also vibrating structures coupled to superconducting electrical devices that are driven by microwave radiation instead of by laser light. The field is motivated both by fundamental questions about quantum mechanics and by more applied aspects, such as the ultrasensitive detection of small displacements or forces and possible uses in quantum information processing.

To enter the quantum regime, mechanical vibrations have to be as cold as possible, which can be achieved by laser cooling. The basic idea is simple enough: send in laser light consisting of photons that do not have quite enough energy to enter the optical cavity, except when they grab an extra quantum of energy from the mechanical vibrations thereby cooling them. The essence of radiation-induced damping of mechanical vibrations was demonstrated³ in 1970 in a macroscopic set-up. In 2004, this principle was first applied to cooling a micro-mechanical resonator using a force created by the thermal effects of light⁴, and in 2006 three groups^{5–7} showed the kind of radiation-pressure laser cooling that has now¹ finally led to cooling a nanomechanical resonator to the quantum ground state. Nevertheless, it proved hard to find a system that combined a sufficiently strong light–mechanics coupling with a weak enough coupling to the thermal

environment, and to pre-cool the system to low temperatures using standard methods.

Chan *et al.*¹ have now overcome these challenges. Their experiment is based on a design introduced two years ago by Painter and colleagues⁸. A silicon nanobeam that has a suitable arrangement of holes (forming a photonic crystal) traps incoming laser light in its central area, in a region not much larger than the wavelength of light, essentially forming an optical cavity (Fig. 1). The beam is free standing, so it can vibrate, and there are standing waves of mechanical vibrations localized at the area where the light is trapped. These are of a high (gigahertz) frequency, making it easier to cool them, and the strong overlap between the tightly localized light field and the mechanical vibrations yields an exceptionally large optomechanical coupling.

In addition, the team exploited the design flexibility of this 'optomechanical crystal' device to engineer a structure in which the damping of vibrational motion is strongly reduced. The combination of all these factors led to successful laser cooling to the ground state: starting with 100 phonons at a temperature of about 20 kelvin, the team¹ was able to reduce the energy of a particular vibrational standing wave to less than one phonon on average. Together with a recent analogous experiment⁹ performed in the microwave domain, Chan and colleagues' study¹ opens the door to exploring the quantum regime of cavity optomechanics.

With the latest advance, it will now become possible to produce non-classical states of light and mechanical motion. One example would be the generation and detection of quantum entanglement in the system — correlations between the light and the mechanical motion that are stronger than anything possible in classical physics. Ultimately, light could even be used to create entanglement between mechanical objects separated by a distance. Another enticing prospect is to engineer optomechanical arrays and circuits that

couple many optical and mechanical oscillations. Such designs could integrate several functionalities, for applications such as sensing and signal processing, or could be used to study the collective dynamics of photons and phonons on a chip. If the coupling between a single photon and a single phonon could be increased 500-fold from the value achieved here, thus making it larger than the photon decay rate in the current set-up¹, then interesting nonlinear quantum effects could be observed.

Finally, researchers in quantum information science are delighted by the prospect of making a device, possibly based on the Painter team's design¹⁰, that converts single phonons to photons. Combining such a device with the already demonstrated¹¹ strong coherent coupling between a nanomechanical resonator and a superconducting two-state quantum system, or qubit, it might be possible to realize an interface between such solid-state qubits and photons, which is much needed for quantum communication applications. ■

Florian Marquardt is at the Institute for Theoretical Physics II, University of Erlangen-Nuremberg, 91058 Erlangen, Germany.
e-mail: florian.marquardt@physik.uni-erlangen.de

1. Chan, J. *et al.* *Nature* **478**, 89–92 (2011).
2. Marquardt, F. & Girvin, S. M. *Physics* **2**, 40 (2009).
3. Braginsky, V. B., Manukin, A. B. & Tikhonov, M. Y. *Sov. Phys. JETP* **31**, 829–830 (1970).
4. Hölcher Metzger, C. & Karrai, K. *Nature* **432**, 1002–1005 (2004).
5. Gigan, S. *et al.* *Nature* **444**, 67–70 (2006).
6. Arcizet, O. *et al.* *Nature* **444**, 71–74 (2006).
7. Schliesser, A. *et al.* *Phys. Rev. Lett.* **97**, 243905 (2006).
8. Eichenfield, M., Chan, J., Camacho, R. M., Vahala, K. J. & Painter, O. *Nature* **462**, 78–82 (2009).
9. Teufel, J. D. *et al.* *Nature* **475**, 359–363 (2011).
10. Safavi-Naeini, A. H. & Painter, O. *N. J. Phys.* **13**, 013017 (2011).
11. O'Connell, A. D. *et al.* *Nature* **464**, 697–703 (2010).

Persistence of soil organic matter as an ecosystem property

Michael W. I. Schmidt^{1*}, Margaret S. Torn^{2,3*}, Samuel Abiven¹, Thorsten Dittmar^{4,5}, Georg Guggenberger⁶, Ivan A. Janssens⁷, Markus Kleber⁸, Ingrid Kögel-Knabner⁹, Johannes Lehmann¹⁰, David A. C. Manning¹¹, Paolo Nannipieri¹², Daniel P. Rasse¹³, Steve Weiner¹⁴ & Susan E. Trumbore¹⁵

Globally, soil organic matter (SOM) contains more than three times as much carbon as either the atmosphere or terrestrial vegetation. Yet it remains largely unknown why some SOM persists for millennia whereas other SOM decomposes readily—and this limits our ability to predict how soils will respond to climate change. Recent analytical and experimental advances have demonstrated that molecular structure alone does not control SOM stability: in fact, environmental and biological controls predominate. Here we propose ways to include this understanding in a new generation of experiments and soil carbon models, thereby improving predictions of the SOM response to global warming.

Understanding soil biogeochemistry is essential to the stewardship of ecosystem services provided by soils, such as soil fertility (for food, fibre and fuel production), water quality, resistance to erosion and climate mitigation through reduced feedbacks to climate change. Soils store at least three times as much carbon (in SOM) as is found in either the atmosphere or in living plants¹. This major pool of organic carbon is sensitive to changes in climate or local environment, but how and on what timescale will it respond to such changes? The feedbacks between soil organic carbon and climate are not fully understood, so we are not fully able to answer these questions^{2–7}, but we can explore them using numerical models of soil-organic-carbon cycling. We can not only simulate feedbacks between climate change and ecosystems, but also evaluate management options and analyse carbon sequestration and biofuel strategies. These models, however, rest on some assumptions that have been challenged and even disproved by recent research arising from new isotopic, spectroscopic and molecular-marker techniques and long-term field experiments.

Here we describe how recent evidence has led to a framework for understanding SOM cycling, and we highlight new approaches that could lead us to a new generation of soil carbon models, which could better reflect observations and inform predictions and policies.

The conundrum of SOM

About a decade ago, a fundamental conundrum was articulated⁸: why, when organic matter is thermodynamically unstable, does it persist in soils, sometimes for thousands of years? Recent advances in physics, material sciences, genomics and computation have enabled a new generation of research on this topic. This in turn has led to a new view of soil-organic-carbon dynamics—that organic matter persists not because of the intrinsic properties of the organic matter itself, but because of physicochemical and biological influences from the surrounding environment that reduce the probability (and therefore rate) of decomposition, thereby allowing the organic matter to persist. In other words, the persistence of soil organic carbon is primarily not a molecular property, but an ecosystem property.

This emerging view has not been fully implemented in global models or research design, for a variety of reasons. First, the knowledge gathered in the past decade has often been published in outlets of traditionally separated disciplines. As a result, confusion has arisen because these different disciplines can use the same vocabulary to mean different things, or vice versa. For example, ‘decomposition rates’ may mean the rate of mass loss of fresh litter, the production rate of CO₂ in a laboratory incubation, or the rate inferred from input and loss of an isotopic tracer present in plant inputs to soil^{9,10}. Second, the complexity of the soil system is difficult to incorporate into one conceptual model or to translate into a tractable yet accurate numerical model. Soil is a realm in which solid, liquid, gas and biology all interact, and the scale of spatial structures spans many orders of magnitude (from nanometre minerals to football-sized soil clods). Indeed, the spatial heterogeneity of biota, environmental conditions and organic matter may have a dominant influence on carbon turnover and trace gas production in soils. Last, the new knowledge remains more qualitative than quantitative. In many cases, it tells us what is important and suggests new model structures, but not how to parameterize them.

Recent insights into carbon cycling

Since pioneering work in the 1980s¹¹, new insights gathered across disciplines (ranging from soil science to marine science, microbiology, material science and archaeology) have challenged several foundational principles of soil biogeochemistry and ecosystem models; in particular, the perceived importance of the ‘recalcitrance’ of the input biomass (the idea that molecular structure alone can create stable organic matter) and of humic substances (biotic or abiotic condensation products). New observations show these to be only marginally important for organic matter cycling^{12,13}. Furthermore, loose use of the term ‘recalcitrance’ has significantly confused the discussion in the past.

We need to ensure that the conceptual framework that supports our understanding of soil carbon cycling is consistent with observations and has a mechanistic basis, as only then can we start to make the necessary

¹Department of Geography, University of Zurich, 8050 Zurich, Switzerland. ²Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ³Energy and Resources Group, University of California, Berkeley, California 94720, USA. ⁴Max Planck Research Group for Marine Geochemistry, University of Oldenburg, Institute for Chemistry and Biology of the Marine Environment, 26129 Oldenburg, Germany. ⁵Max Planck Research Group for Marine Microbiology, Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany. ⁶Institute of Soil Science, Leibniz Universität Hannover, 30419 Hannover, Germany. ⁷Department of Biology, University of Antwerp, 2610 Wilrijk, Belgium. ⁸Department of Crop and Soil Science, Oregon State University, Corvallis, Oregon 97331, USA. ⁹Lehrstuhl für Bodenkunde, Technische Universität München, 85354 Freising, Germany. ¹⁰Department of Crop and Soil Sciences, Atkinson Center for a Sustainable Future, Cornell University, Ithaca, New York 14853, USA. ¹¹School of Civil Engineering and Geosciences, Institute for Research on Environment and Sustainability, Newcastle University, Newcastle NE1 7RU, UK. ¹²Department of Plant, Soil and Environmental Sciences, University of Firenze, 50144 Firenze, Italy. ¹³Norwegian Institute for Agricultural and Environmental Research, 1432 Ås, Norway. ¹⁴Structural Biology, Weizmann Institute, 76100 Rehovot, Israel. ¹⁵Max Planck Institute for Biogeochemistry, 07745 Jena, Germany.

*These authors contributed equally to this work.

advances in terrestrial ecology and improve our ability to predict soil responses to changes in climate, vegetation or management. Here we articulate key insights into soil carbon cycling synthesized from research of the past decade, and describe the research challenges they pose for the coming decade.

Molecular structure and decomposition

The initial decomposition rate of plant residues correlates broadly with indices of their bulk chemical composition, such as the nitrogen content or the fraction of plant residue that cannot be solubilized by strong acid treatments (often operationally defined as 'lignin')¹⁴. Accordingly, the molecular structure of biomass and organic material has long been thought to determine long-term decomposition rates in the mineral soil. However, using compound-specific isotopic analysis, molecules predicted to persist in soils (such as lignins or plant lipids) have been shown to turn over more rapidly than the bulk of the organic matter (Fig. 1)^{12,15–17}. Furthermore, other potentially labile compounds, such as sugars, can persist not for weeks but for decades. We therefore cannot extrapolate the initial stages of litter decomposition to explain the persistence of organic compounds in soils for centuries to millennia—other mechanisms protect against decomposition. Perhaps certain compounds require co-metabolism with another (missing) compound, or microenvironmental conditions restrict the access (or activity) of decomposer enzymes (for example, hydrophobicity, soil acidity, or sorption to surfaces¹⁸).

Soil humic substances

The prevalence of humic substances in soil has been assumed for decades¹⁹. Previous generations of soil chemists relied on alkali and acid extraction methods²⁰ and observations of the extracted (or residual)

functional-group chemistry to describe the presence of operationally defined 'humic and fulvic acids' and 'humin'. Humic substances were thought to comprise large, complex macromolecules that were the largest and most stable SOM fraction. However, we now understand that these components represent only a small fraction of total organic matter^{13,21–23}; direct, *in situ* observations, rather than verifying the existence of these large, complex molecules, in fact find smaller, simpler molecular structures, as visualized in Fig. 2 (refs 13, 22, 23). Some of what is extracted as humic acids may be fire-derived^{24,25}, although these compounds are rare in soil without substantial fire-derived organic matter. In any case, there is not enough evidence to support the hypothesis that the *de novo* formation of humic polymers is quantitatively relevant for humus formation in soils.

Fire-derived organic matter

Fire-derived organic matter (also called char, black carbon or pyrolysed carbon) is found in many soils, sediments and water bodies, and can comprise up to 40% of total SOM in grasslands and boreal forests²⁶. It is not inert, but its decomposition pathways remain a mystery. Fire-derived carbon was suspected to be more stable in soil than other organic matter because of its fused aromatic ring structures and the old radiocarbon ages of fire residues isolated from soil²⁷. However, fire-derived carbon does undergo oxidation and transport, as we now know from archaeological settings²⁸, soils^{29,30}, and from breakdown products in river³¹ and ocean water^{32,33}. In a field experiment, fire-derived residues were even observed to decompose faster than the remaining bulk organic matter, with 25% lost over 100 years (ref. 29). Spectroscopic characterization shows that combustion temperature affects the degree of aromaticity and the size of aromatic sheets, which in turn determine short-term mineralization

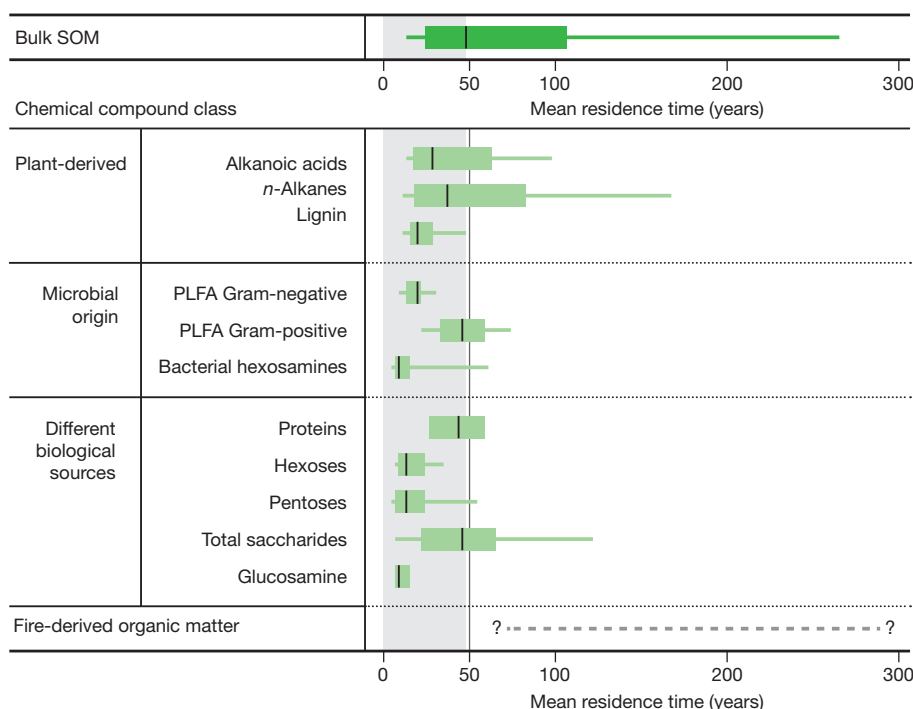


Figure 1 | Molecular structure does not control long-term decomposition of soil organic matter (SOM). Certain plant-derived molecules (classically, long-chain alkanolic acids, *n*-alkanes, lignin and other structural tissues) often persist longer than others while plant biomass is decaying. In mineral soil, however, these relatively persistent components appear to turn over faster than the bulk soil (top row), except for fire-derived organic matter (bottom row). Even components that appear chemically labile, including proteins and saccharides of plant and microbial origin ('Different biological sources'), instead seem to turn over (on average) at rates similar to those of bulk SOM, that is, on the order of years or even decades. Thus, over time, the importance of initial quality fades and the initially fast-cycling compounds are just as likely to persist as the

slow^{12,15}. This figure compiles data from surface horizons of 20 long-term field experiments (up to 23 years) in temperate climate, using ¹³C labelling to trace the residence time of bulk SOM and of individual molecular compounds. The variation in turnover time is also seen in the compounds of microbial origin analysed for ¹³C content, phospholipid fatty acids (PLFA) produced by Gram-negative and Gram-positive bacteria and amino sugars (hexosamines). Redrawn from ref. 15 (with permission); for clarity, we have excluded outliers, and we have added the tentative data on fire-derived organic matter. Data points: thin horizontal lines, 10th and 90th percentiles; box, 25th and 75th percentiles; central vertical line, median.

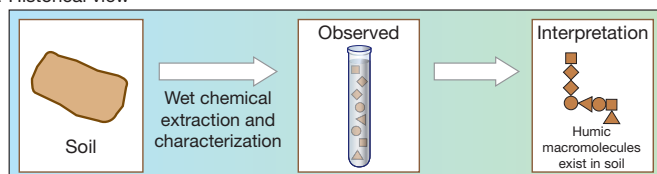
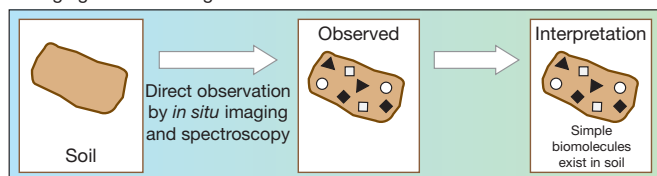
a Historical view**b Emerging understanding**

Figure 2 | In soil, the existence of humic substances has not been verified by direct measurements. **a**, Based on chemical analysis of the extracted materials (Observed), the *de novo* formation of humic polymers (Interpretation) was postulated to be an important source of recalcitrant SOM. **b**, Direct high-resolution *in situ* observations with non-destructive techniques (Observations) have been able to explain the functional group chemistry of the extracted humic substances as relatively simple biomolecules (Interpretation), without the need to invoke the presence of unexplainable macromolecules¹⁰⁰. Moreover, the chemical mixture of SOM is spatially distinct on a nanometre scale, and the aromatic/carboxylate-rich compounds characteristic of the bulk extracted humic substances have not been found *in situ* even when looking at the submicrometre scale (using near-edge X-ray fine structure spectroscopy combined with scanning transmission X-ray microscopy)²².

rates^{22,34–36}. To reconcile the observations of decomposability with the old radiocarbon ages of fire-derived carbon deposits^{37,38}, it has been suggested that physical protection and interactions with soil minerals play a significant part in black-carbon stability over long periods of time³⁹.

Influence of roots

Root-derived carbon is retained in soils much more efficiently than are above-ground inputs of leaves and needles^{40–42}. Isotopic analyses and comparisons of root and shoot biomarkers confirm the dominance of root-derived molecular structures in soil⁴³ and of root-derived carbon in soil microorganisms⁴⁴. Preferential retention of root-derived carbon has been observed in temperate forests^{45,46}, for example, where below-ground inputs, including fungal mycelia, make up a bigger fraction of new carbon in SOM than do leaf litter inputs^{44,47}. In addition to many above-ground inputs being mineralized in the litter layer, root and mycorrhizal inputs have more opportunity for physico-chemical interactions with soil particles⁴⁰. At the same time, fresh root inputs may ‘prime’ microbial activity, leading to faster decomposition of older organic matter^{48,49} as well as changing community composition⁵⁰. Carbon allocation by plants thus plays an important part in soil carbon dynamics, but it is not known how future changes in plant allocation will affect soil carbon stocks⁵¹.

Physical disconnection

The soil volume occupied by micro-organisms is considerably less than 1%: this occupied volume is distributed heterogeneously in small-scale habitats, connected by water-saturated or unsaturated pore space¹⁸. The availability of spatially and temporally diverse habitats probably gives rise to the biodiversity that we see in soil, but this fragmentation of habitat may restrict carbon turnover. At present, we are far from being able to quantify the complex processes of soil structure development and fragmentation, which have different space scales and timescales depending on soil type, texture and management¹⁰. The physical disconnection between decomposer and organic matter is likely to be one reason for persistence of deep SOM. The specific pedological processes operating in a given soil type that influence the distribution of organisms and substrates, such as bioturbation and formation of preferential flow paths, need to be taken into account to understand and quantify subsoil carbon dynamics, and thus its vulnerability to decomposition⁵².

Deep soil carbon

There is a lot more deep soil carbon than we once thought, and the underlying processes inhibiting its turnover are still largely unknown. Despite their low carbon concentrations, subsoil horizons contribute to more than half of the global soil carbon stocks⁵³. In fact, the response of deep soils to land-use change can equal that from the top 30 cm of soil, even though typically only the shallow depths are explicitly represented in models⁵⁴. Inputs of carbon to the subsoil include dissolved organic matter, root products, and transported particulates from the surface⁵⁵, but the relative importance of different sources is not known⁵⁶. Based on depth trends of elemental composition (decreasing C/N ratio), isotopic composition (increasing $\delta^{13}\text{C}$ values) and individual organic compounds, microbial products make up more organic matter in subsoil horizons than do plant compounds⁵⁷.

Organic matter in subsoil horizons is characterized by very long turnover times that increase with depth—radiocarbon ages of 1,000 to >10,000 years are common—but the reasons for this are not clear. Microbial activity may be reduced by suboptimal environmental conditions, nutrient limitation or energy scarcity, and organic matter may be less accessible because of its sparse density or association with reactive mineral surfaces. Microbial biomass decreases with soil depth⁵⁸, and community composition changes to reflect an increase in substrate specialization⁵⁹. Recent studies suggest that energy limitation, or the converse—‘priming’ (see below) by root exudates or dissolved organic carbon—is an important factor in the subsurface^{48,49}. Most studies concerning these factors, however, have been conducted in the laboratory, and their relevance *in situ* needs evaluation. If we do not understand these mechanisms of stabilization, we cannot predict the vulnerability of deep SOM to change.

Thawing permafrost

Permafrost soils store as much carbon (up to $1,672 \times 10^{15}$ g; ref. 60) as was believed a decade ago to exist in all soils worldwide. During permafrost thaw, which is expected to become widespread owing to climate change, much of this SOM may be vulnerable to rapid mineralization⁶¹ if it is primarily stabilized by freezing temperatures⁶². There is evidence that old carbon is mobilized following permafrost thaw^{61,63}, which indicates that organic matter previously locked in the permafrost is highly vulnerable. Moreover, the accelerated decomposition may increase nitrogen availability, which would amplify the direct effects of warming on microbial activity. Alleviation of nitrogen limitation in tundra experiments led to large and rapid carbon losses, including older carbon^{64,65}. Over the very long term, however, formation of pedogenic reactive minerals in former permafrost soils may act to stabilize SOM^{66,67}, and development of soil structure may lead to physical disconnection between organic matter and decomposers.

Despite some important recent research, surprisingly little is known about permafrost biogeochemistry and how the landscape would evolve with warming. Key questions surround the extent to which permafrost carbon is additionally stabilized by other processes beyond freezing, and the extent to which the active layer becomes saturated and anaerobic. The extent, rates and spatial variability of these processes are still largely unknown for permafrost soils, forming one of the major uncertainties in predicting climate–carbon feedbacks.

Soil micro-organisms

Soil microbial diversity and activity can be characterized at molecular resolution, but the quantitative linkages to ecosystem function are uncertain^{68,69}. Soil micro-organisms influence SOM cycling not only via decomposition but also because microbial products are themselves important components of SOM⁷⁰. As a result, environmental change can influence soil carbon cycling through changes in both metabolic activity and community structure. For example, microbial community shifts following nitrogen additions can have large effects on decomposition rates^{50,51}. New genetic and protein-based tools enable the quantification of soil microbiological abundance and functioning (for example, enzymatic gene expression), and can describe the microbial community com-

position with very high taxonomic resolution⁷¹. Nevertheless, the challenge remains of synthesizing this immense amount of detailed information⁷² and linking it to the rates and routes of SOM processing. To quantitatively relate microbial genomics to ecosystem function, we need a better understanding of microbial functional redundancy.

Implications of these insights

Taken together, these eight insights paint a broad picture of carbon cycling in soil that has implications for fundamental research, land management, and climate change prediction and mitigation (Fig. 3). They suggest that the molecular structure of plant inputs and organic matter has a secondary role in determining carbon residence times over decades to millennia, and that carbon stability instead mainly depends on its biotic and abiotic environment (it is an ecosystem property). Most soil carbon derives from below-ground inputs and is transformed, through oxidation by microorganisms, into the substances found in the soil. By moving on from the concept of recalcitrance and making better use of the breadth of relevant research, the emerging conceptual model of soil organic carbon cycling will help to unravel the mysteries surrounding the fate of plant- and fire-derived inputs and how their dynamics vary between sites and soil depths, and to understand feedbacks to climate change. We argue that the persistence of organic matter in soil is largely due to complex interactions between organic matter and its environment, such as the interdependence of compound chemistry, reactive mineral surfaces, climate, water availability, soil acidity, soil redox state and the presence of potential degraders in the immediate microenvironment. This does not mean that compound chemistry is not important for decomposition rates, just that its influence depends on environmental factors. Rather than describing organic matter by decay rate, pool, stability or level of 'recalcitrance'—as if these were properties of the compounds themselves—organic matter should be described by quantifiable environmental characteristics governing stabilization, such as solubility, molecular size and functionalization⁷³.

Soil response to global environmental change

We now consider how these insights affect our use of numerical models. Such models are powerful tools for quantifying the complex interactions

and feedbacks that will underpin soil responses to global change. A variety of models that include SOM dynamics have informed our response to environmental issues, including agricultural management, bioremediation and environmental water research⁷⁴. Most model testing, however, has been at local-to-regional spatial scales, spanning seasons to decades (although the century-long Rothamsted experiments are a noteworthy exception). In the long term or at a global scale, mechanisms of SOM stabilization and destabilization that are not currently embedded in models have the potential to dominate soil carbon dynamics, making it vital that models are correct for the right reasons. Recent model intercomparisons reveal large differences among predictions of soil carbon stocks and fluxes in the next century, for example², demonstrating how sensitive global carbon cycling is to assumptions about SOM decomposition dynamics.

Recent advances in our mechanistic understanding of soils, such as those described above, have not yet been incorporated into the widely used models of SOM cycling, which are all structured around the idea that a type, or pool, of organic material will have an intrinsic decay rate^{75–78}. These models rely on simple proxies—such as soil texture as a surrogate for sorption and other organo-mineral interactions, and litter quality (such as lignin:N ratios or structural carbon groupings) as a means of partitioning plant inputs into pools of different turnover times—but in general these parameters are not consistent with the observations that are starting to emerge. Global models largely ignore deep mineral soils and are only now beginning to address the accumulation and loss of carbon in peatlands and permafrost⁷⁹. Even more importantly, parameterizations based on litter chemistry may correlate well to initial rates of litter decomposition, but they have little relationship to the rates of decomposition for microbial residues or to organic matter sorbed to mineral surfaces or isolated in aggregates. Moreover, most models that make any allowances for microbial biomass treat it as a pool of carbon, rather than as an agent that affects the decomposition rate of SOM. The large disagreement among predictions of soil carbon fluxes in a warmer world highlights both the complexity of the many potential feedbacks to climate and the uncertainty that arises as a result².

How does the perspective that SOM persistence is an ecosystem property inform our understanding of the response of decomposition

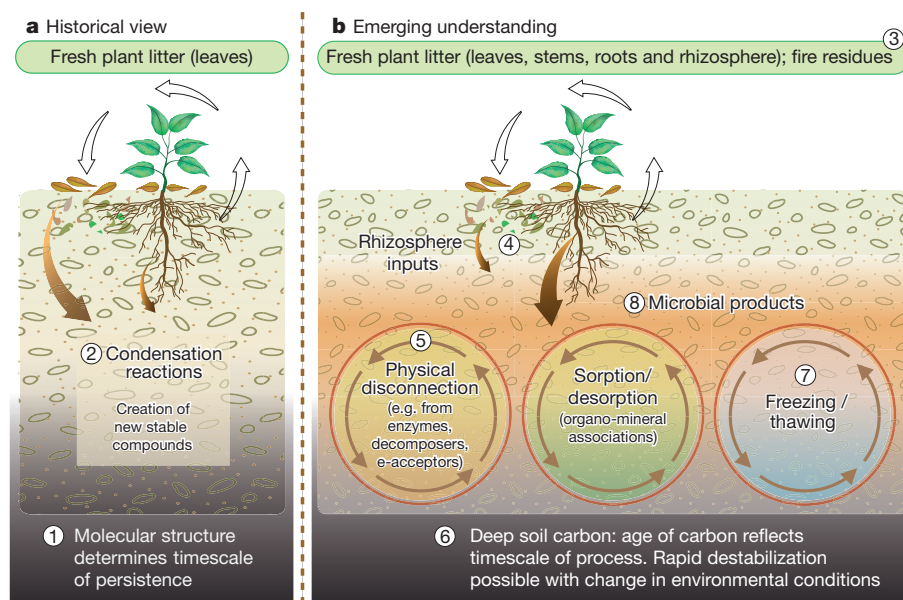


Figure 3 | A synopsis of all eight insights, contrasting historical and emerging views of soil carbon cycling. The historical view (a) has emphasized above-ground plant carbon inputs and organic matter in the top 30 cm of soil. Stable organic matter is seen to comprise mainly selectively preserved plant inputs and *de novo* synthesis products like humic substances, whose chemical complexity and composition render them nearly inert relative to microbial degradation. The emerging understanding (b) is that the molecular structure of organic material does not necessarily determine its stability in soil (1; molecular

structure). Rather, SOM cycling is governed by multiple processes (5) shaped by environmental conditions (such as physical heterogeneity). Plant roots and rhizosphere inputs (4; roots) make a large contribution to SOM, which is mainly partial degradation and microbial products and fire residues (3) rather than humic substances (2). The vulnerability of deep soil carbon (6; deep carbon) to microbial degradation (8; soil micro-organisms) in a changing environment, such as thawing permafrost (7; thawing permafrost) remains a key uncertainty.

to warming? The conventional assumption that older SOM is recalcitrant implies that this large carbon pool is highly temperature sensitive, because Arrhenius kinetics tells us that reactions with higher activation energies are more temperature-sensitive than those with low activation energies^{80–82}. Our ecosystem perspective suggests that the mechanisms governing the timing and magnitude of a response to a change in temperature are far more complex than this, as further physical, chemical and biological mechanisms controlling decomposition and stabilization would also be affected^{81,83,84}. A recent incubation study of soils from a wide range of sites found that lower initial decomposition rates were associated with higher temperature sensitivity but not with any change in SOM quality indices⁸⁵, suggesting that multiple stabilization mechanisms are temperature sensitive. Nevertheless, it is not yet possible to predict the integrated response of decomposition to changes in climate. In fact, we could use the ability to accurately predict temperature response as a guide to the degree of mechanistic representation that we need in our next generation of soil carbon models.

Phyto-engineering

Phyto-engineering to produce plant tissues high in chemical compounds resistant to rapid mineralization, such as plant lipids and lignin, has been suggested as a means to increase carbon sequestration⁸⁶. This strategy is called into question, however, if the molecular structure of plant compounds does not determine stability on the timescales necessary for significant carbon sequestration^{12,87}. More generally, sequestration strategies based on adding recalcitrant material to soils, whether through plant selection for recalcitrant tissues or through biochar amendments, must be re-evaluated. Enhancing root carbon input to soils might be a more promising avenue, but it is not known what root properties influence rhizodeposition rates or stability⁴³, or the extent to which root inputs will stimulate (prime) decomposition of other SOM.

Biochar

Biochar (intentionally pyrolysed biomass) has gained much attention in recent years as a means to increase soil fertility and store carbon in soil for decades to centuries⁸⁸. However, certain types of biochar can degrade relatively rapidly in some soils, probably depending on the conditions under which they were produced, which suggests that pyrolysis could be optimized to generate a more stable biochar. But as with natural fire residues, persistence over the long term may also be affected by interactions with minerals and by soil conditions (for microorganisms capable of char oxidation and for abiotic oxidation). Whether interactions of fire-derived carbon with soil minerals may be manipulated to enhance stability, and what the trade-offs might be with fertility benefits, are not known. Biochar is likely to be a useful part of sequestration-mitigation strategies, but more understanding of the variation in its decay rates is needed before we can develop simple (that is, policy-relevant) quantitative relationships between biochar additions and expected sequestration.

Vulnerability of soil to degradation

The vulnerability of SOM to degradation will depend on the nature of the disturbance as well as the stabilization and destabilization mechanisms at play in a given ecosystem. Hence, as with carbon stability, the vulnerability of soil stocks should not be assessed according to the classes of organic matter present, but rather according to the mechanisms through which organic matter is stabilized or made assimilable in that soil, and how these interacting physical, chemical and biological factors respond to change^{5,6}. Improved understanding of SOM destabilization is needed to enhance efforts to avoid soil degradation and accelerate recovery of degraded soils.

The way forward

Soils are now in the 'front line' of global environmental change—we need to be able to predict how they will respond to changing climate, vegetation, erosion and pollution so that we can better understand their role in the Earth system and ensure that they continue to provide for humanity⁸⁹ and the natural world. The conceptual framework of soil

carbon cycling presented here, that residence time is a property of the interactions between organic matter and the surrounding soil ecosystem, will help us get nearer to these goals⁷. This will require developed and entirely new lines of research and modelling, including: (1) applying a new generation of field experiments and analytical tools to study the processes driving SOM stabilization and destabilization; (2) developing a new generation of soil biogeochemistry models that represent the mechanisms driving soil response to global change; and (3) joining forces and connecting the disparate research communities that are studying, managing and predicting SOM cycling and terrestrial ecology.

The next generation of experiments

Although not a novel recommendation, we cannot overstate the need for long-term, manipulative experiments designed to test soil-based hypotheses. In some countries, long-term ecological observational networks already exist, but most were designed with vegetation or hydrology goals. Many are in danger of being discontinued. Although preserving these experiments is crucial, they may not be sufficient to untangle individual soil processes. In the near term, new disciplines and techniques could be applied to ongoing experiments, allowing the investigation of changes that occurred after decades of manipulation⁹⁰. Focus is needed on long-term, controlled manipulations of entire soil profiles (that is, to a metre or more depth) to investigate distinct mechanisms *in situ*, and on observatories allowing quantification of budgets, such as large-scale lysimeters. In addition, research approaches are needed that combine manipulations with spatial gradients—and thus timescales—for variables and processes of interest. These new experiments should be designed to help determine the key soil functional traits for understanding and modelling thresholds in SOM storage and loss. Such traits, including soil depth, mineral charge density and pH, vary spatially, but we suggest that their spatial distributions are ultimately predictable according to geologic setting, disturbance and management history, climate and ecosystem plant characteristics—in other words, the six state factors: climate, organisms, relief, parent material, time and human activity⁹¹. One of the major weaknesses of current models is the lack of representation of edaphic characteristics (that is, those physical and chemical features that are intrinsic to the soil)—and the fact that the major stabilization mechanisms will vary spatially with soil type and topographic positions.

Tracing pathways, fluxes and biology

When combined with manipulative experiments, new analytical techniques and instrumentation to study elements, isotopes and molecules in terrestrial ecosystems offer great potential for revealing the mechanisms underpinning soil carbon stability. Advances in physics, material sciences, genomics and computation continue to create new research opportunities. Because many, if not most, organic molecules in soils are of microbial origin, experiments are needed that identify the long-term drivers of microbial-cell and microbial-product decomposition, rather than focusing on the immediate fate of fresh plant material.

We propose extending the systems biology approach to the non-living environment that surrounds organisms. Individual molecules could then be traced back from the soil into the cell via metabolic pathways and specific gene expressions. As with medicine, where structure–function relationships led to the development of genomics and proteomics, allowing illness to be treated before symptoms appear, we foresee that the integration of molecular, biological and physical information will provide soil science with a more mechanistic basis for predictions.

However, a major obstacle remains. The molecular complexity of SOM is extraordinary, and the metabolic products of higher plants and the diverse soil microbial community are mixed together in a three-dimensional inorganic soil matrix. An essential step to overcoming this obstacle is the identification of intact molecular structures in soils. In marine and freshwater sciences, ultrahigh-resolution (Fourier-transform ion cyclotron resonance) mass spectrometry has made it possible to identify tens of thousands of organic compounds in water and soil water, a major step towards implementing a systems biology

approach in soils³². Ultrahigh-resolution mass spectrometry could be applied to soils by combining it with ionization techniques such as matrix-assisted laser desorption ionization.

For soils, any high-resolution molecular technique will have to be combined with imaging techniques in order to address spatial relationships and heterogeneity. Moreover, new spatial techniques will help uncover how microbial community structure, enzyme production and decomposition activity are influenced by environmental conditions and plant processes. For example, combining fluorescence *in situ* hybridization, which produces a spatial map of the microbial community, with secondary ion mass spectrometry imaging on a nanometre scale, can be a powerful way to link biota with processes at the submicrometre scale^{92,93}.

Beyond imaging, new methods to trace particle and solute transport (for example, viral DNA labels) can help us to understand the processes linking deep and surface soils, and isotopic advances reveal both the movement and the chemical transformation of carbon in soil. The value of isotopically labelled inputs has been greatly amplified by new tools that allow precise measurements on small samples: it is now possible to follow labelled elements in the environment (for example, ¹⁴C), and to 'fingerprint' specific plant compounds and microbial products in soil, and therefore to determine how decomposition pathways and substrate ages interact. For measuring carbon isotopic values in gas and water fluxes, we no longer need to rely on weekly measurements, which carry the danger of missing episodic but crucial events. Quantum cascade and cavity ring-down laser spectroscopy allow high-frequency observations of the molecular and isotopic composition of soil gas, efflux and water.

Integrative computational databases

Advanced analytical methods are generating an ever-increasing amount of data on SOM. Soil DNA databases are currently being developed nationally⁹⁴ and internationally (for example, soil microbial genomics libraries). In parallel, data-fusion-type techniques are now being applied to the large data sets generated by microbial fingerprinting and spectroscopic methods⁹⁵. However, we are still lacking the integrative databases and computational tools that would enable us to identify significant relationships between the detailed molecular composition of soil organic carbon, expressions of microbial and plant activities, and soil environmental conditions. The development of international libraries and high-performance computational databases for molecular SOM research will be necessary to take advantage of new molecular tools and create synthesis across analytical platforms.

Mechanisms in global ecosystem models

The current soil models embedded in Earth system models are structured around 3–5 pools of organic substrates, with transformation rates modified by empirical correlations to soil temperature and water content, and with clay content as a proxy for mineral stabilization of organic matter^{75–78}. These models have proven extremely useful, but in the case of long-term feedbacks to climate, a more realistic representation of what governs organic-matter stability will be needed to more accurately inform predictions and policies. Table 1 characterizes what the most widely used ecosystem models do, relative to the insights described above, and gives our recommendations for how to incorporate recent advances into models. Many of the recommendations in Table 1 are what we consider 'low-hanging fruit': that is, modifications that could be made in existing model frameworks, with existing knowledge and data, and that should make significant improvements.

Specifically, we suggest changes along three 'axes'. First, decay of organic matter is a biological process and should be treated as such in models. For example, changes in below-ground carbon allocation or in nutrient availability may alter microbial community composition and activity, which in turn will alter the rate of degradation and the types of organic matter that are degraded. Rhizospheric inputs of easily assimilable substrates may aid in, or prime, the decomposition of compounds that would otherwise be selectively avoided by microorganisms⁴⁹. This breadth of biological influence is not currently accounted for.

Second, evidence demonstrating the relatively fast degradation of char and lignin implies that all substrates are degradable within a suitable environment. Likewise, molecules that chemically resemble one another can exhibit very different residence times, depending on whether or not they are protected from decay. The way forward for global land models is to change their organizing principle from carbon pools with intrinsic decomposition rates (based on correlations with texture or litter quality, and modified by climate and land-use type) to more mechanistic representations of the stabilization processes that actually govern carbon dynamics and therefore the strength of climate feedbacks.

Third, representing fine-scale processes and heterogeneity at the large scale of global models is a major challenge for the field. Box models, which have been the basis for modelling soil carbon, assume a mean behaviour at a specific spatial scale. Frequently, however, the distributions of substrates, microorganisms and environmental conditions are highly skewed. For example, a soil that is on average at optimal soil water content for microbial activity will contain pockets that are too wet (such

Table 1 | Representation of soil carbon in ecosystem models and recommendations for potential improvements

Insight	Properties of most published models	Recommendations
1. Molecular structure	Decay rate of all pools keyed to substrate (or texture in CENTURY-type models*) and modified by moisture and temperature as constant Q_{10} above 0 °C.	Model decay rate as function of substrate properties and positions in microenvironment, microbial activity, and soil conditions including pH, temperature and moisture. See 4, 5, 6, 8.
2. Humic substances	Have a cascade of increasing intrinsic recalcitrance due to decomposition and synthesis.	Replace the cascade with cycling of organic matter into and out of microbial biomass. See 1, 8.
3. Fire-derived carbon	Do not include fire-residues as inputs or SOM. Do not represent decay of analogous substrates.	Add input pathway for fire-derived carbon. Add aromatic compounds to SOM types.
4. Roots	Parameterize litter quality with leaf/needle chemistry. Have simplified root and dissolved organic carbon inputs.	Use separate characterizations for below-ground and above-ground inputs. See 6.
5. Physical heterogeneity	Lack physical processes, such as aggregation (some have tillage factor), spatial heterogeneity, or processes that would produce priming effect†.	Non-normal probability distributions, density-dependent terms for organic matter and microbial biomass. Parameters from 3D, fine-resolution models.
6. Soil depth	No change in processes or rate constants with depth of soil or carbon input. Site-level tuning required to reproduce long turnover times.	Representations of mineral associations, root and dissolved organic inputs, and physical disconnections. Explicit depth resolution for decomposition and transport.
7. Permafrost	Lack processes governing permafrost soil carbon cycling. Lack fully coupled methane biogeochemistry.	Add O ₂ limitation and freezing effects on CO ₂ and CH ₄ production. Develop soil columns to represent inundation, permafrost thaw and thermokarst.
8. Soil micro-organisms	Treat microbial biomass as pool of active carbon. Lack effects of microbial community or enzymes on rates and decomposition products.	Create and model microbial functional types, analogous to plant functional types. Introduce full soil nitrogen cycle coupled to carbon cycle.

Shown are properties of the soil carbon component of published ecosystem models used for global change and carbon cycle analysis, and recommendations for potential improvements, grouped according to the eight insights developed in this Perspective (see 'Recent insights into carbon cycling' section). Globally implemented land models, such as Orchidee, LPJ, IBIS, CASA* and CLM, are based on CENTURY, CN, or RothC soil models^{75–78}.

* Some models use texture (clay content) to determine the amount of carbon in the slowest-cycling pool. However, soils with the same texture differ twofold in carbon stock and turnover time owing to differences in mineralogy, for example⁹⁹. One improvement would be to replace texture with reactive iron and aluminium or mineral surface charge density, estimated globally from a pedotransfer function.

† Priming effect means that carbon input rate has positive effect on decomposition rate.

as inside aggregates) as well as too dry (such as inside hydrophobic microsities). At the landscape scale, a small inundated area exerts a disproportionate effect on average methane emissions of a global model grid cell. Because decomposition responds nonlinearly to its drivers, average environmental conditions are imprecise predictors of the perturbed system, and models that can represent this spatial dimension, from the micro-site to the profile to the landscape, are likely to fare better than current models. Advances in numerical methods make possible more mechanistic treatments of transport, taxa-specific microbial requirements and other three-dimensional phenomena⁹⁶.

Moving forward requires identifying which parameters are critical, developing practical representations of these processes and parameters, and testing predictions of SOM dynamics against observations at the relevant scales ('ground-truthing'). One reason that models have not incorporated recent scientific developments is the lack of appropriate data for parameterization and testing. The advanced analytical techniques and long-term experiments called for above are vital to fill this gap. In addition, more effort is needed both by modellers and empirical scientists to facilitate model evaluation. For example, the ¹⁴C content of respired CO₂ or leached dissolved organic carbon is a powerful constraint on underlying mechanisms of changes in stocks. In addition, ¹⁴C 'clocks' the time carbon has spent in the ecosystem, and is the only way to quantify carbon residence time in undisturbed systems. Yet most ecosystem models lack a ¹⁴C tracer that would allow it to be used for testing. Clearly, a comprehensive database of ¹⁴C measurements is needed.

Join forces and connect research communities

The cycling of organic matter is the subject of many different disciplines—from marine chemistry^{32,97} to low-temperature geology to archaeology⁹. Even within SOM research, there have been at least two separate and disconnected directions of research⁹⁸. There are those studying litter decomposition, with a focus on the biotic breakdown of plant inputs, often in forest organic litter layers or agricultural systems. On the other hand, there are those focused on organo-mineral interactions in the mineral soil⁹⁸. Other examples of separated research approaches include agronomic versus ecological questions, aquatic versus terrestrial environments, and laboratory versus field-based experiments. Cross-fertilization is especially needed between empirical scientists and modellers in the context of global change. Insights into mechanisms and observational data will improve predictions, and, in return, the needs of models will motivate useful experiments.

More generally, though, the major advances in our understanding of soils will come from research grounded in the theory of many disciplines and in the practice of many approaches. The future research agenda for soils will integrate many different fields and have broader goals than it might have had in the past, with longer time horizons, wider spatial coverage, and an imperative to connect carbon, water and nutrient cycles, so as to understand the soil–plant system as a crucial part of our biosphere.

1. Fischlin, A. *et al.* in *Climate Change 2007: Impacts, Adaptation and Vulnerability* (eds Parry, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. & Hanson, C. E.) 211–272 (Cambridge Univ. Press, 2007).
2. Friedlingstein, P. *et al.* Climate-carbon cycle feedback analysis: results from the C⁴MIP model intercomparison. *J. Clim.* **19**, 3337–3353 (2006).
A systematic comparison of model predictions of soil carbon response to climate change.
3. von Lützow, M. & Kögel-Knabner, I. Temperature sensitivity of soil organic matter decomposition—what do we know? *Biol. Fertil. Soils* **46**, 1–15 (2009).
A review and outline of research needs about the response of soil organic matter to rising temperatures
4. Kirschbaum, M. U. F. The temperature dependence of organic-matter decomposition — still a topic of debate. *Soil Biol. Biochem.* **38**, 2510–2518 (2006).
5. Heimann, M. & Reichstein, M. Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature* **451**, 289–292 (2008).
6. Trumbore, S. E. & Czimczik, C. I. An uncertain future for soil carbon. *Science* **321**, 1455–1456 (2008).
7. Sollins, P., Homann, P. & Caldwell, B. A. Stabilization and destabilization of soil organic matter: mechanisms and controls. *Geoderma* **74**, 65–105 (1996).
Described mechanisms of SOM stabilization involving environmental controls.

8. Hedges, J. I. *et al.* The molecularly-uncharacterized component of nonliving organic matter in natural environments. *Org. Geochem.* **31**, 945–958 (2000).
Formulated the fundamental question of why, when organic matter is thermodynamically unstable, does it persist in soils, sometimes for thousands of years?
9. Hedges, J. I. & Oades, J. M. Comparative organic geochemistries of soils and sediments. *Org. Geochem.* **27**, 319–361 (1997).
10. Totsche, K. U. *et al.* Biogeochemical interfaces in soil: the interdisciplinary challenge for soil science. *J. Plant Nutr. Soil Sci.* **173**, 88–99 (2010).
11. Oades, J. M. The retention of organic matter in soils. *Biogeochemistry* **5**, 35–70 (1988).
12. Marschner, B. *et al.* How relevant is recalcitrance for the stabilization of organic matter in soils? *J. Plant Nutr. Soil Sci.* **171**, 91–110 (2008).
13. Kleber, M. & Johnson, M. G. Advances in understanding the molecular structure of soil organic matter: implications for interactions in the environment. *Adv. Agron.* **106**, 77–142 (2010).
14. Melillo, J. M., Aber, J. D. & Muratore, J. F. Nitrogen and lignin control of hardwood leaf litter decomposition dynamics. *Ecology* **63**, 621–626 (1982).
15. Amelung, W., Brodowski, S., Sandhage-Hofmann, A. & Bol, R. Combining biomarker with stable isotope analysis for assessing the transformation and turnover of soil organic matter. *Adv. Agron.* **100**, 155–250 (2008).
A review including a compilation of the surprisingly rapid and overlapping turnover times of individual molecular compounds previously suspected to have 'slow' turnover.
16. Knorr, M., Frey, S. D. & Curtis, P. S. Nitrogen additions and litter decomposition: a meta-analysis. *Ecology* **86**, 3252–3257 (2005).
17. Grandy, A. S. & Neff, J. C. Molecular C dynamics downstream: the biochemical decomposition sequence and its impact on soil organic matter structure and function. *Sci. Total Environ.* **404**, 297–307 (2008).
18. Ekschmitt, K. *et al.* Soil-carbon preservation through habitat constraints and biological limitations on decomposer activity. *J. Plant Nutr. Soil Sci.* **171**, 27–35 (2008).
19. Stevenson, F. J. *Humus Chemistry* (Wiley, 1994).
20. Olk, D. C. & Gregorich, E. G. Overview of the symposium proceedings, "Meaningful pools in determining soil carbon and nitrogen dynamics". *Soil Sci. Soc. Am. J.* **70**, 967–974 (2006).
21. von Lützow, M. *et al.* Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions — a review. *Eur. J. Soil Sci.* **57**, 426–445 (2006).
22. Lehmann, J. *et al.* Spatial complexity of soil organic matter forms at nanometre scales. *Nature Geosci.* **1**, 238–242 (2008).
23. Sutton, R. & Spósito, G. Molecular structure in soil humic substances: the new view. *Environ. Sci. Technol.* **39**, 9009–9015 (2005).
24. Haumaier, L. & Zech, W. Black carbon — possible source of highly aromatic components of soil humic acids. *Org. Geochem.* **23**, 191–196 (1995).
25. Trompowsky, P. M. *et al.* Characterization of humic like substances obtained by chemical oxidation of eucalyptus charcoal. *Org. Geochem.* **36**, 1480–1489 (2005).
26. Preston, C. M. & Schmidt, M. W. I. Black (pyrogenic) carbon: a synthesis of current knowledge and uncertainties with special consideration of boreal regions. *Biogeosciences* **3**, 397–420 (2006).
A summary of the current understanding of the formation, properties and fate of fire-residues in natural ecosystems.
27. Schmidt, M. W. I. & Noack, A. G. Black carbon in soils and sediments: analysis, distribution, implications, and current challenges. *Glob. Biogeochem. Cycles* **14**, 777–794 (2000).
28. Cohen-Ofri, I., Weiner, L., Boaretto, E., Mintz, G. & Weiner, S. Modern and fossil charcoal: aspects of structure and diagenesis. *J. Archaeol. Sci.* **33**, 428–439 (2006).
29. Hammes, K., Torn, M. S., Lapenas, A. G. & Schmidt, M. W. I. Centennial black carbon turnover observed in a Russian steppe soil. *Biogeosciences* **5**, 1339–1350 (2008).
30. Major, J., Lehmann, J., Rondon, M. & Goode, C. Fate of soil-applied black carbon: downward migration, leaching and soil respiration. *Glob. Change Biol.* **16**, 1366–1379 (2010).
31. Kim, S., Kaplan, L. A., Benner, R. & Hatcher, P. G. Hydrogen-deficient molecules in natural riverine water samples — evidence for the existence of black carbon in DOM. *Mar. Chem.* **92**, 225–234 (2004).
32. Dittmar, T. & Paeng, J. A heat-induced molecular signature in marine dissolved organic matter. *Nature Geosci.* **2**, 175–179 (2009).
33. Ziolkowski, L. A. & Druffel, E. R. M. Aged black carbon identified in marine dissolved organic carbon. *Geophys. Res. Lett.* **37**, L16601 (2010).
34. Nguyen, B. T., Lehmann, J., Hockaday, W. C., Joseph, S. & Masiello, C. A. Temperature sensitivity of black carbon decomposition and oxidation. *Environ. Sci. Technol.* **44**, 3324–3331 (2010).
35. Keiluweit, M., Nico, P. S., Johnson, M. G. & Kleber, M. Dynamic molecular structure of plant biomass-derived black carbon (biochar). *Environ. Sci. Technol.* **44**, 1247–1253 (2010).
36. Liang, B. *et al.* Stability of biomass-derived black carbon in soils. *Geochim. Cosmochim. Acta* **72**, 6069–6078 (2008).
37. Cheng, C. H. & Lehmann, J. Ageing of black carbon along a temperature gradient. *Chemosphere* **75**, 1021–1027 (2009).
38. Lehmann, J. *et al.* Australian climate-carbon cycle feedback reduced by soil black carbon. *Nature Geosci.* **1**, 832–835 (2008).
39. Brodowski, S., John, B., Flessa, H. & Amelung, W. Aggregate-occluded black carbon in soil. *Eur. J. Soil Sci.* **57**, 539–546 (2006).
40. Rasse, D. P., Rumpel, C. & Dignac, M. F. Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation. *Plant Soil* **269**, 341–356 (2005).

41. Kong, A. Y. Y. & Six, J. Tracing root vs. residue carbon into soils from conventional and alternative cropping systems. *Soil Sci. Soc. Am. J.* **74**, 1201–1210 (2010).
42. Balesdent, J. & Balabane, M. Major contribution of roots to soil carbon storage inferred from maize cultivated soils. *Soil Biol. Biochem.* **28**, 1261–1263 (1996).
43. Mendez-Millan, M., Dignac, M. F., Rumpel, C., Rasse, D. P. & Derenne, S. Molecular dynamics of shoot vs. root biomarkers in an agricultural soil estimated by natural abundance ^{13}C labelling. *Soil Biol. Biochem.* **42**, 169–177 (2010).
44. Kramer, C. et al. Recent (4 year old) leaf litter is not a major source of microbial carbon in a temperate forest mineral soil. *Soil Biol. Biochem.* **42**, 1028–1037 (2010).
45. Bird, J. A., Kleber, M. & Torn, M. S. ^{13}C and ^{15}N stabilization dynamics in soil organic matter fractions during needle and fine root decomposition. *Org. Geochem.* **39**, 465–477 (2008).
46. Bird, J. A. & Torn, M. S. Fine roots vs. needles: A comparison of ^{13}C and ^{15}N dynamics in a ponderosa pine forest soil. *Biogeochemistry* **79**, 361–382 (2006).
47. Godbold, D. L. et al. Mycorrhizal hyphal turnover as a dominant process for carbon input into soil organic matter. *Plant Soil* **281**, 15–24 (2006).
48. Fontaine, S. et al. Stability of organic carbon in deep soil layers controlled by fresh carbon supply. *Nature* **450**, 277–280 (2007).
49. Kuzyakov, Y. Priming effects: interactions between living and dead organic matter. *Soil Biol. Biochem.* **42**, 1363–1371 (2010).
50. Ågren, G. I., Bosatta, E. & Magill, A. H. Combining theory and experiment to understand effects of inorganic nitrogen on litter decomposition. *Oecologia* **128**, 94–98 (2001).
51. Janssens, I. A. et al. Reduction of forest soil respiration in response to nitrogen deposition. *Nature Geosci.* **3**, 315–322 (2010).
52. Chabbi, A., Kögel-Knabner, I. & Rumpel, C. Stabilised carbon in subsoil horizons is located in spatially distinct parts of the soil profile. *Soil Biol. Biochem.* **41**, 256–261 (2009).
53. Jobbágy, E. G. & Jackson, R. B. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecol. Appl.* **10**, 423–436 (2000).
54. Trumbore, S. E., Davidson, E. A., de Camargo, P. B., Nepstad, D. C. & Martinelli, L. A. Belowground cycling of carbon in forests and pastures of Eastern Amazonia. *Glob. Biogeochem. Cycles* **9**, 515–528 (1995).
55. Rumpel, C. & Kögel-Knabner, I. Deep soil organic matter — a key but poorly understood component of terrestrial C cycle. *Plant Soil* **338**, 143–158 (2011).
- A comprehensive overview of key challenges to quantitative understanding of deep soil carbon.**
56. Kalbitz, K., Schwesig, D., Rethemeyer, J. & Matzner, E. Stabilization of dissolved organic matter by sorption to the mineral soil. *Soil Biol. Biochem.* **37**, 1319–1331 (2005).
57. Torn, M. S. et al. Organic carbon and carbon isotopes in modern and 100-year-old soil archives of the Russian steppe. *Glob. Change Biol.* **8**, 941–953 (2002).
58. Fierer, N., Allen, A. S., Schimel, D. P. & Holden, P. A. Controls on microbial CO_2 production: a comparison of surface and subsurface soil horizons. *Glob. Change Biol.* **9**, 1322–1332 (2003).
59. Kramer, C. & Gleixner, G. Soil organic matter in soil depth profiles: distinct carbon preferences of microbial groups during carbon transformation. *Soil Biol. Biochem.* **40**, 425–433 (2008).
60. Tarnocai, C. et al. Soil organic carbon pools in the northern circumpolar permafrost region. *Glob. Biogeochem. Cycles* **23**, GB2023, doi:10.1029/2008GB003327 (2009).
61. Schuur, E. A. G. et al. The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. *Nature* **459**, 556–559 (2009).
62. Schuur, E. A. G. et al. Vulnerability of permafrost carbon to climate change: implications for the global carbon cycle. *Bioscience* **58**, 701–714 (2008).
63. Nowinski, N. S., Taneva, L., Trumbore, S. E. & Welker, J. M. Decomposition of old organic matter as a result of deeper active layers in a snow depth manipulation experiment. *Oecologia* **163**, 785–792 (2010).
64. Mack, M. C., Schuur, E. A. G., Bret-Harte, M. S., Shaver, G. R. & Chapin, F. S. Ecosystem carbon storage in arctic tundra reduced by long-term nutrient fertilization. *Nature* **431**, 440–443 (2004).
65. Nowinski, N. S., Trumbore, S. E., Schuur, E. A. G., Mack, M. C. & Shaver, G. R. Nutrient addition prompts rapid destabilization of organic matter in an arctic tundra ecosystem. *Ecosystems* **11**, 16–25 (2008).
66. Striegl, R. G., Aiken, G. R., Dornblaser, M. M., Raymond, P. A. & Wickland, K. P. A decrease in discharge-normalized DOC export by the Yukon River during summer through autumn. *Geophys. Res. Lett.* **32**, L21413, doi:10.1029/2005GL024413 (2005).
67. Kawahigashi, M., Kaiser, K., Rodionov, A. & Guggenberger, G. Sorption of dissolved organic matter by mineral soils of the Siberian forest tundra. *Glob. Change Biol.* **12**, 1868–1877 (2006).
68. Raes, J. & Bork, P. Molecular eco-systems biology: towards an understanding of community function. *Nature Rev. Microbiol.* **6**, 693–699 (2008).
69. Morales, S. E. & Holben, W. E. Linking bacterial identities and ecosystem processes: can 'omic' analyses be more than the sum of their parts? *FEMS Microbiol. Ecol.* **75**, 2–16 (2011).
70. Kögel-Knabner, I. The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. *Soil Biol. Biochem.* **34**, 139–162 (2002).
71. McGuire, K. L. & Treseder, K. K. Microbial communities and their relevance for ecosystem models: decomposition as a case study. *Soil Biol. Biochem.* **42**, 529–535 (2010).
72. von Mering, C. et al. Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315**, 1126–1130 (2007).
73. Kleber, M. What is recalcitrant soil organic matter? *Environ. Chem.* **7**, 320–332 (2010).
74. Manzoni, S. & Porporato, A. Soil carbon and nitrogen mineralization: theory and models across scales. *Soil Biol. Biochem.* **41**, 1355–1379 (2009).
75. Kucharik, C. J. et al. Measurements and modeling of carbon and nitrogen cycling in agroecosystems of southern Wisconsin: potential for SOC sequestration during the next 50 years. *Ecosystems* **4**, 237–258 (2001).
76. Jenkinson, D. S. The turnover of organic carbon and nitrogen in soil. *Phil. Trans. R. Soc. Lond.* **329**, 361–368 (1990).
77. Parton, W. J., Ojima, D. S., Cole, C. V. & Schimel, D. S. in *Quantitative Modeling of Soil Forming Processes* (eds Bryant, R. B. & Arnold, R. W.) 147–167 (Special Publication, Soil Science Society of America, 1994).
78. Thornton, P. E. & Rosenbloom, N. A. Ecosystem model spin-up: estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model. *Ecol. Model.* **189**, 25–48 (2005).
79. Khvorostyanov, D. V., Krinner, G., Ciais, P., Heimann, M. & Zimov, S. A. Vulnerability of permafrost carbon to global warming. Part I: model description and role of heat generated by organic matter decomposition. *Tellus B* **60**, 250–264 (2008).
80. Arrhenius, S. Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren. *Z. Phys. Chem.* **4**, 226–248 (1889).
81. Davidson, E. A. & Janssens, I. A. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature* **440**, 165–173 (2006).
82. Knorr, W., Prentice, I. C., House, J. I. & Holland, E. A. Long-term sensitivity of soil carbon turnover to warming. *Nature* **433**, 298–301 (2005).
83. Kirschbaum, M. U. F. The temperature dependence of organic matter decomposition: seasonal temperature variations turn a sharp short-term temperature response into a more moderate annually averaged response. *Glob. Change Biol.* **16**, 2117–2129 (2010).
84. Fang, C., Smith, P., Moncrieff, J. B. & Smith, J. U. Similar response of labile and resistant soil organic matter pools to changes in temperature. *Nature* **433**, 57–59 (2005).
85. Craine, J. M., Fierer, N. & McLaughlin, K. K. Widespread coupling between the rate and temperature sensitivity of organic matter decay. *Nature Geosci.* **3**, 854–857 (2010).
86. Lorenz, K., Lal, R., Preston, C. M. & Nierop, K. G. J. Strengthening the soil organic carbon pool by increasing contributions from recalcitrant aliphatic bio(macro)molecules. *Geoderma* **142**, 1–10 (2007).
87. Thevenot, M., Dignac, M. F. & Rumpel, C. Fate of lignins in soils: a review. *Soil Biol. Biochem.* **42**, 1200–1211 (2010).
88. Lehmann, J. A handful of carbon. *Nature* **447**, 143–144 (2007).
89. Sachs, J. et al. Monitoring the world's agriculture. *Nature* **466**, 558–560 (2010).
90. Richter, D. D., Hofmockel, M., Callahan, M. A., Powlson, D. S. & Smith, P. Long-term soil experiments: keys to managing Earth's rapidly changing ecosystems. *Soil Sci. Soc. Am. J.* **71**, 266–279 (2007).
91. Amundson, R. & Jenny, H. The place of humans in the state factor theory of ecosystems and their soils. *Soil Sci.* **151**, 99–109 (1991).
92. Amstalden van Hove, E. R., Smith, D. F. & Heeren, R. M. A. A concise review of mass spectrometry imaging. *J. Chromatogr. A* **1217**, 3946–3954 (2010).
93. Herrmann, A. M. et al. Nano-scale secondary ion mass spectrometry — a new analytical tool in biogeochemistry and soil ecology: A review article. *Soil Biol. Biochem.* **39**, 1835–1850 (2007).
94. Ranjard, L. et al. Biogeography of soil microbial communities: a review and a description of the ongoing French national initiative. *Agron. Sustain. Dev.* **30**, 359–365 (2010).
95. Pascault, N. et al. In situ dynamics of microbial communities during decomposition of wheat, rape and alfalfa residues. *Microb. Ecol.* **60**, 816–828 (2010).
96. Xu, T. F. Incorporating aqueous reaction kinetics and biodegradation into TOUGHREACT: applying a multiregion model to hydrobiogeochemical transport of denitrification and sulfate reduction. *Vadose Zone J.* **7**, 305–315 (2008).
97. Jiao, N. et al. Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nature Rev. Microbiol.* **8**, 593–599 (2010).
98. Sollins, P., Swanston, C. & Kramer, M. Stabilization and destabilization of soil organic matter — a new focus. *Biogeochemistry* **85**, 1–7 (2007).
99. Torn, M. S., Trumbore, S. E., Chadwick, O. A., Vitousek, P. M. & Hendricks, D. M. Mineral control of soil organic carbon storage and turnover. *Nature* **389**, 170–173 (1997).
100. Kelleher, B. P. & Simpson, A. J. Humic substances in soils: are they really chemically distinct? *Environ. Sci. Technol.* **40**, 4605–4611 (2006).

Acknowledgements The European Science Foundation Network MOLTER sponsored the workshop at which the idea for this Perspective was conceived. Support for M.W.I.S. and M.S.T. was also provided by the US Department of Energy (contract no. DE-AC02-05CH11231).

Author Contributions M.W.I.S. coordinated the MOLTER-sponsored workshop mentioned above; the ideas were developed by all authors. M.W.I.S. and M.S.T. participated actively and equally in the writing of the manuscript and the drafting of the figures. All authors provided input into the drafting and the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence should be addressed to M.W.I.S. (Michael.Schmidt@geo.uzh.ch) or M.S.T. (mstorn@lbl.gov).

Deep sequencing reveals 50 novel genes for recessive cognitive disorders

Hossein Najmabadi^{1,2}, Hao Hu^{3*}, Masoud Garshasbi^{1,3*}, Tomasz Zemojtel⁴, Seyede Sedigheh Abedini¹, Wei Chen^{3,5}, Masoumeh Hosseini¹, Farkhondeh Behjati¹, Stefan Haas⁴, Payman Jamali⁶, Agnes Zecha³, Marzieh Mohseni¹, Lucia Püttmann³, Leyla Nouri Vahid¹, Corinna Jensen³, Lia Abbasi Moheb^{1,3}, Melanie Bienek³, Farzaneh Larti¹, Ines Mueller³, Robert Weissmann³, Hossein Darvish¹, Klaus Wrogemann^{3,7}, Valeh Hadavi², Bettina Lipkowitz³, Sahar Esmaeeli-Nieh³, Dagmar Wiczorek⁸, Roxana Kariminejad², Saghar Ghasemi Firouzabadi¹, Monika Cohen⁹, Zohreh Fattahi¹, Imma Rost¹⁰, Faezeh Mojahedi¹¹, Christoph Hertzberg¹², Atefeh Dehghan¹³, Anna Rajab¹⁴, Mohammad Javad Soltani Banavandi¹, Julia Hoffer³, Masoumeh Falah¹, Luciana Musante³, Vera Kalscheuer³, Reinhard Ullmann³, Andreas Walter Kuss^{3†}, Andreas Tzschach³, Kimia Kahrizi¹ & H. Hilger Ropers³

Common diseases are often complex because they are genetically heterogeneous, with many different genetic defects giving rise to clinically indistinguishable phenotypes. This has been amply documented for early-onset cognitive impairment, or intellectual disability, one of the most complex disorders known and a very important health care problem worldwide. More than 90 different gene defects have been identified for X-chromosome-linked intellectual disability alone, but research into the more frequent autosomal forms of intellectual disability is still in its infancy. To expedite the molecular elucidation of autosomal-recessive intellectual disability, we have now performed homozygosity mapping, exon enrichment and next-generation sequencing in 136 consanguineous families with autosomal-recessive intellectual disability from Iran and elsewhere. This study, the largest published so far, has revealed additional mutations in 23 genes previously implicated in intellectual disability or related neurological disorders, as well as single, probably disease-causing variants in 50 novel candidate genes. Proteins encoded by several of these genes interact directly with products of known intellectual disability genes, and many are involved in fundamental cellular processes such as transcription and translation, cell-cycle control, energy metabolism and fatty-acid synthesis, which seem to be pivotal for normal brain development and function.

Early-onset cognitive impairment, or intellectual disability, is an unresolved health care problem and an enormous socio-economic burden. Most severe forms of intellectual disability are due to chromosomal abnormalities or defects in specific genes. For many years, research into the genetic causes of intellectual disability and related disorders has focused on X-chromosome-linked intellectual disability (XLID). It has become clear, however, that X-linked forms account for only 10% of intellectual disability cases, which means that the vast majority of the underlying genetic defects must be autosomal¹. For severe forms of intellectual disability, autosomal-dominant inheritance is rare because most affected individuals do not reproduce, but recent observations suggest that in outbred Caucasian populations, a significant portion of the sporadic cases may be due to dominant *de novo* mutations^{2–4}. So far, relatively little is known about the role of autosomal recessive intellectual disability (ARID), because in Western societies, where most of the research takes place, its investigation has been hampered by infrequent parental consanguinity and small family sizes.

In most Northern African countries, and also in the Near and Middle East, parental consanguinity and large families are common; for example, in Iran, 40% of the families are consanguineous and about two-thirds of the population is 30 years of age or younger.

Since 2004, we have performed systematic array-based consanguinity mapping in 272 consanguineous Iranian families. In several dozen families, we have defined single linkage intervals and mapped the underlying gene defects^{5,6}, and by subsequent mutation screening of candidate genes from these intervals, we and others identified several novel ARID genes (for review see refs 1, 7).

Recently, exome enrichment and next-generation sequencing have been introduced as a cost-effective and fast strategy for comprehensive mutation screening and disease-gene identification in the coding portion of the human genome^{8–10}. To unravel the molecular basis of ARID in a systematic fashion, we have now used a related, but more targeted, approach. Instead of sequencing entire exomes in consanguineous families, we have focused on the exons from homozygous linkage intervals known to carry the genetic defect. Before sequencing, these exons were enriched by hybrid capture using custom-made oligonucleotide arrays as baits. All patients had cognitive impairment (mostly moderate or severe, see Supplementary Table 1), and in a subset of the families there were signs of autism spectrum disorder. More information about the families and their clinical features, quality controls performed to validate the sequence variants observed and to assess their pathogenicity, as well as other methodological details are provided in Supplementary Information.

¹Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, 19857 Tehran, Iran. ²Kariminejad-Najmabadi Pathology & Genetics Centre, 14667 Tehran, Iran. ³Department Human Molecular Genetics, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. ⁴Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. ⁵Max-Planck-Centrum für Molekulare Medizin, 13092 Berlin, Germany. ⁶Shahroud Welfare Organization, 36156 Semnan, Iran. ⁷Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Manitoba R3E0J9, Canada. ⁸Institut fuer Humangenetik, Universitaetsklinikum, 45122 Essen, Germany. ⁹Kinderzentrum Muenchen, 81377 Muenchen, Germany. ¹⁰Zentrum fuer Humangenetik und Laboratoriumsmedizin Dr Klein und Dr Rost, 82152 Martinsried, Germany. ¹¹Mashhad Medical Genetic Counseling Center, 91767 Mashhad, Iran. ¹²Kinderneurologie und Sozialpaediatric, Vivantes-Klinikum Neukölln, 12351 Berlin, Germany. ¹³Yazd Welfare Organization, 89178 Yazd, Iran. ¹⁴Genetics Unit, Ministry of Health, Directorate General of Health Affairs, Royal Hospital, Muscat 113, Oman. [†]Present address: Institute for Human Genetics, University Medicine Greifswald & Interfaculty Institute for Genetics and Functional Genomics, Ernst Moritz Arndt University, 17489 Greifswald, Germany.

*These authors contributed equally to this work.

Mutations in known and novel intellectual disability genes

In 115 out of 136 families studied, plausible causal defects were observed, and in 78 of these, a single, apparently disease-causing mutation could be identified (see Supplementary Fig. 1, Tables 1 and 2 and Supplementary Table 2). Twenty-eight protein-truncating changes were found, including frameshift, splice-site and nonsense mutations, as well as whole-exon deletions, plus several smaller in-frame deletions of varying size. In 26 families listed in Table 1, we identified known, mostly syndromic forms of ARID, including rare metabolic defects and storage disorders, such as an atypical form of Tay–Sachs' disease and Sanfilippo's syndrome (mucopolysaccharidosis IIIB), as well as intellectual disability with congenital abnormalities, such as a Joubert-like syndrome resulting from *AHI1* mutations, observed in two unrelated families. Two families were also found with allelic *PRKCG* mutations, implicated previously in spinocerebellar ataxia, and two families carried different allelic mutations in the *SRD5A3* gene, associated with Kahrizi's syndrome, a recently elucidated congenital glycosylation disorder^{11,12}.

Two mutations involving the adaptor protein complex 4 were observed, namely in the *AP4M1* and *AP4E1* genes, which encode different AP-4 subunits. AP-4 is involved in the recognition and sorting of cargo protein transported from the trans-Golgi network to the endosomal-lysosomal system. Another possibly pathogenic change was found in the *AP4B1* gene, but its effect may be obscured by a *PEX6* mutation in the same family, which causes a severe peroxisome biosynthesis disorder¹³ and probably accounts for most of the clinical features. In highly inbred families, coexistence of two different recessive defects is not unexpected and is the most plausible explanation for the complex phenotypes in at least two families with novel forms of ARID (M154 and M189, see Table 2).

Mutations in the *SLC2A1* gene, which encodes a glucose transporter, the *PRKRA* gene with a role in dysautonomia, and the *MED13L* gene, previously associated with intellectual disability and cardiac symptoms, were the only plausible causes of intellectual disability in three families with non-syndromic intellectual disability. None of the respective families showed signs of dysautonomia or cardiac abnormalities. In all other families, the phenotype was characteristic for the molecular defect, including family M198 with folate receptor deficiency, a rare syndromic form of ARID that can often be

treated by oral administration of folinic acid¹⁴. Further details are provided in Table 1.

Apparently pathogenic changes were also found in 50 genes that had not been previously implicated in ARID (see Table 2). Thirty of the relevant families had non-syndromic forms of intellectual disability, whereas 22 exhibited syndromic forms. Only two of the novel ARID genes were mutated in more than a single family. Two different missense mutations with high pathogenicity scores were detected in *ZNF526*, which encodes a krüppel-type zinc-finger protein. One of these changes was observed in DNA samples collected from two distinct families with non-syndromic intellectual disability, but closer inspection revealed that these families, which live in the same city in the northwestern part of Iran, share a common haplotype and thus must be distantly related. In these families, no other potentially disease-causing and co-segregating change could be identified. Zinc-finger proteins are transcriptional regulators, and other krüppel-type zinc-finger genes have been implicated in intellectual disability before¹⁵. Recent protein interaction studies have indicated a role for *ZNF526* in promoting messenger RNA translation and cell growth (N. Hubner *et al.*, personal communication). Another gene within which disease-causing mutations were found in two families was *ELP2*. It encodes a subunit of the RNA polymerase II elongator complex, which is a histone acetyltransferase component of RNA polymerase II. This gene is involved in the acetylation of histones H3 and probably H4, and it may have a role in chromatin remodelling.

Mutations affecting housekeeping genes

In the *LARP7* gene, we found a frameshift mutation in a family with intellectual disability and microcephaly. *LARP7* is a negative transcriptional regulator of polymerase II genes, acting by means of the 7SK RNP system. Within the 7SK RNP complex, the positive transcription elongation factor b (P-TEFb) is sequestered in an inactive form, preventing RNA polymerase II phosphorylation and subsequent transcriptional elongation. Hitherto, no disease association has been reported for *LARP7*.

Presumably causative homozygous mutations were also found in *KDM5A* and *KDM6B*. These genes encode histone demethylases that specifically demethylate histone H3 at lysine 4 and lysine 27, respectively, and they both have a central role in the histone code. We have

Table 1 | Mutations identified in known genes for intellectual disability or related disorders

Family	Gene	Mutation	LOD score	Length (Mb)	OMIM no.	Diagnosis, clinical features
8500306	<i>AHI1</i>	R329X	2.65	10.35	608629	Joubert's syndrome 3
M332	<i>AHI1</i>	R495H	3.2	11.1	608629	Joubert's syndrome 3
M254	<i>AP4E1</i>	V454fs	2.5	13.57	607244	Microcephaly, paraplegia
M004	<i>AP4M1</i>	E193K	1.9	16.75	602296	Microcephaly, paraplegia
M324	<i>BBS7</i>	533del2aa	3.24	8.2	209900	Bardet–Biedl's syndrome
M107	<i>CA8</i>	R237Q	2.4	4.02	613227	Ataxia, cerebellar hypoplasia
M175	<i>COL18A1</i>	L1587fs	2.1	9.8	267750	Knobloch's syndrome (eye and brain development)
G026	<i>FAM126A</i>	Splice site*	2.4	15.46	610532	Hypomyelination-cataract
M198	<i>FOLR1</i>	Splice site*	2.1	16.95	136430	Folate receptor deficiency
M165	<i>HEXA</i>	C58Y	2.7	15.91	272800	Psychomotor delay, mild Tay–Sachs' disease
8600276†	<i>L2HGDH</i>	R335X	5.1	13.39	609584	Hydroxyglutaric aciduria
M142	<i>MED13L</i>	R1416H	1.9	9.17	608808	Non-syndromic ID, no cardiac involvement
8600486	<i>NAGLU</i>	R565Q	2.8	13.25	252920	Sanfilippo's syndrome, MPS IIIB
8500234	<i>PDHX</i>	R15H	3.13	35.17	245349	Pyruvate dehydrogenase defect
M331	<i>PEX6</i>	L534P	3.8	10.83	601498	Peroxisome biogenesis disorder
8307998	<i>PMM2</i>	Y106F	2.67	6.71	212065	Glycosylation disorder CDG Ia
8600273	<i>PRKCG</i>	V177fs	2.53	0.72	605361	Spinocerebellar ataxia 14
M146	<i>PRKCG</i>	D480Y	2.1	7.45	605361	Spinocerebellar ataxia 14
8600162	<i>PRKRA</i>	S235T	2.1	40.02	612067	Non-syndromic ID
8600042	<i>SLC2A1</i>	V237M	3.73	16.7	606777	Non-syndromic ID
8700017	<i>SRD5A3</i>	Y169C	4.8	10.5	612713	Kahrizi's syndrome, CDG
M069†	<i>SRD5A3</i>	A68fs	3.01	10.44	612713	Kahrizi's syndrome, CDG
G008	<i>SURF1</i>	W227R	1.8	4.59	185620	Leigh's syndrome, very mild form
8600041	<i>TH</i>	R202H	2.1	7.23	605407	infantile parkinsonism, Segawa's syndrome
M017N	<i>VRK1</i>	R133C	3.4	3	607596	Pontocerebellar hypoplasia
M196	<i>WDR62</i>	G705G	2.1	18.33	600176	Microcephaly, cerebellar atrophy

CDG, congenital disorder of glycosylation; fs, frameshift; ID, intellectual disability; LOD, logarithm of the odds; MPS, mucopolysaccharidosis; OMIM, Online Mendelian Inheritance in Man.

* See Supplementary Information for further details.

† Remotely related, degree of consanguinity is not clear, analysis performed under conservative assumption of second degree consanguinity.

previously shown that mutations in another lysine-specific histone demethylase, *KDM5C* (also called *JARID1C*), are a relatively frequent cause of X-linked intellectual disability¹⁶. In two other families, we observed apparently pathogenic mutations that involved histones directly: a frameshift mutation in the *HIST1H4B* gene which belongs to the histone 4 family, and a *HIST3H3* missense mutation with high pathogenicity scores that was the only plausible change in a family with non-syndromic intellectual disability. Together, at least ten of the novel candidate genes for ARID involve histone structure, histone modification, chromatin remodelling or the regulation of transcription, and many of these genes are functionally linked to known and novel intellectual disability genes, as shown in Fig. 1a.

Several other mutated genes are directly or indirectly involved in the regulation of translation. A homozygous frameshift mutation inactivating the *TRMT1* gene was detected in a family with non-syndromic intellectual disability. *TRMT1* is an RNA methyltransferase that dimethylates a single guanine residue at position 26 of most tRNAs. Previously we and others have shown that inactivation of the X-linked gene *FTSJ1*, another RNA methyltransferase, also gives rise to non-syndromic intellectual disability^{17,18}, and we have recently identified several ARID families with truncating mutations in a third RNA methyltransferase (L.A.M. *et al.*, manuscript in preparation). A large deletion in the *EEF1B2* gene was the only detectable defect in another family with non-syndromic intellectual disability. *EEF1B2* encodes the elongation factor 1 β , which is involved in the transport of aminoacyl-tRNAs to the ribosomes. In yet another family with non-syndromic intellectual disability, a missense change was found in *ADRA2B*. This gene encodes a brain-expressed G-protein-coupled receptor that associates with EIF2B, a guanine exchange factor regulating translation¹⁹; notably, *ADRA2B* also interacts with the 14-3-3 protein, which in turn associates with RGS7, another novel ARID gene product that regulates G-protein signalling. Finally, in a family with a syndromic form of intellectual disability, a missense change was found in the *POLR3B* gene, involving a nucleotide with a very high conservation score and predicted to be pathogenic by Mutation Taster²⁰. *POLR3B* encodes the second-largest core component of RNA polymerase III, which synthesizes small RNAs such as tRNAs and 5S rRNAs²¹ and also interacts with ENTPD1, the product of a novel candidate gene for intellectual disability (see GeneCards, <http://www.genecards.org/cgi-bin/cardsearch.pl?search=POLR3B> and Table 2). Together, these observations indicate that gene defects interfering with transcription and translation are particularly important causes of intellectual disability.

However, we also found pathogenic mutations affecting other fundamental cellular functions and pathways such as cell-cycle control, as illustrated by a mutation inactivating *CCNA2*, and another one truncating *SCAPER*, a specific regulator of the CCNA2-CDK2 complex (see Fig. 1b). The *C11orf46* gene encodes TTI2, a subunit of the Triple T complex, which is required for the establishment of cell-cycle checkpoints and for DNA-damage signalling²². Other mutations involved fatty-acid synthesis and turnover (*ACBD6*, *FASN* and *PECR*; see Table 2), protein degradation (*UBR7*), splicing (*ZCCHC8*) and cell migration (*LAMA1*).

Intellectual disability genes with brain-specific functions

Not surprisingly, several mutations involved genes with neuron- or brain-specific functions. For example, we found a frameshift mutation abolishing the function of *CACNA1G*, a T-type calcium channel with a critical role in the generation of GABA_B receptor-mediated spike and wave discharges in the thalamocortical pathway^{23,24}. A nonsense mutation inactivated *ZBTB40*, which has a role in glia cell differentiation²⁵, and other observed changes are expected to interfere with the regulation of neurotransmission, exocytosis or neurotransmitter release. Our study also adds several novel intellectual-disability-associated genes to the Ras and Rho pathway (see Fig. 1c); for example, a convincing missense mutation in the *RALGDS* gene was the only variant detected in one family with non-syndromic intellectual disability. This gene encodes an effector of the Ras-related GTPase Ral, which stimulates the dissociation of GDP from the Ras-related RalA and RalB GTPases, thereby allowing GTP binding and activation of the GTPases²⁶. Regulators of small GTPases were among the first genes to be implicated in non-syndromic intellectual disability^{27,28}. We also found a homozygous frameshift mutation in *CNKSRI*, which is physically associated with *RALGDS*. Homozygous carriers of this mutation have a severe syndromic phenotype with quadrupedal gait. *CNKSRI* binds to rhophilin (Online Mendelian Inheritance in Man (OMIM) 601031), a Rho effector, suggesting that it acts as a scaffold protein and mediates crosstalk between the Ras and Rho GTPase signalling pathways²⁹. Neither *RALGDS* nor *CNKSRI* had been implicated in intellectual disability so far; thus, both are novel ARID genes.

Genes without obvious link to intellectual disability

For several of the sequence variants, there is no obvious functional link between the molecular defect and intellectual disability. This applies to *LINS1* and *NDST1*, and it is not easy to understand why

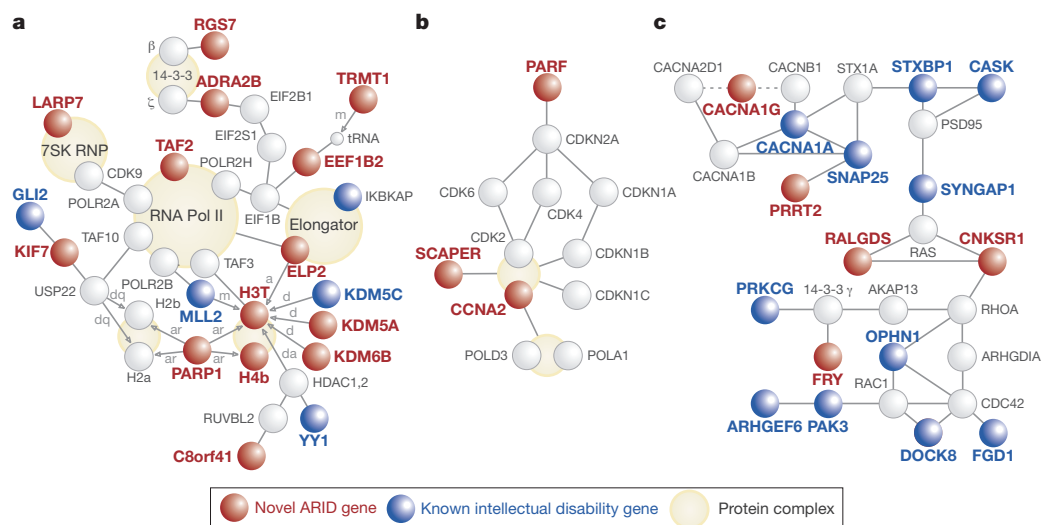


Figure 1 | Known and novel intellectual disability genes form protein and regulatory networks. **a**, Transcriptional/translational network. **b**, Cell-cycle-related network. **c**, Ras/Rho/PSD95 network. Connecting edges in the figure stand for protein–protein interactions. Arrows define direction of post-translational protein modifications: a, acetylation; ar, ADP-ribosylation; d,

demethylation; da, deacetylation; dq, deubiquitination; m, methylation. Dotted lines indicate modulation of gene function. Data were obtained in part by using the INGENUITY software package (<http://www.ingenuity.com>) and by literature mining. More details about these proteins and their interactions are provided in Table 2 and in Supplementary Information.

Table 2 | Apparently causative variants in novel (candidate) genes for intellectual disability

Family	Phenotype	Gene	Mutation	LOD score	Length (Mb)	Supporting evidence
M008† M173	S NS, ASD	<i>ACBD6</i> <i>ADK</i>	G22fs H324R	2.65 5.1	6.46 9.68	P; binds long-chain acyl-CoA molecules, role in fatty acid synthesis or turnover ⁴⁴ . S, P; only change in family. Adenosine kinase, regulates adenosine levels in the brain. Overexpression leads to learning impairment in mice ⁴⁵ ; knockout mice develop lethal neonatal liver steatosis ³⁰ . In human, a different gene defect has been found in this condition.
M266-2	NS	<i>ADRA2B</i>	R440G	2.53	24.97	S, P; GPCR regulating adrenergic neurons in the CNS. Associates with EIF2B, a GEF regulating translation ¹⁹ . Also associates with 14-3-3, which interacts with RGS7, mutated in family 8700136.
M226	NS	<i>ASCC3</i>	S1564P	3.2	62.80	S, P, E; helicase that is part of the activating signal co-integrator complex, enhances NF-κB and AP1. Interacts with RARS2, implicated in pontocerebellar hypoplasia 6 ⁴⁶ .
M007L‡	NS	<i>ASCL1</i>	A41S	2.4	18.13	Encodes the bHLH factor MASH1, critical role in neuronal commitment and differentiation ^{47,48} .
M182 G001	NS NS	<i>C11orf46</i> <i>C12orf57</i>	R236H M1V	2.1 2.5	12.39 11.19	P, E; encodes subunit of the Triple T complex, role in regulation of DNA damage response ²² . S; function hitherto unknown. May overlap neighbouring <i>ANT1</i> (DRPLA) gene (see UCSC Genome Browser, hg18; OMIM 125370).
M100	NS	<i>C8orf41</i>	P367L	3.3	6.44	S, P, E; C8orf41 associates with RUVBL2 ⁴⁹ , which is involved in regulation of transcription and interacts with HDACs ⁵⁰ .
G015	NS	<i>C9orf86</i>	A562P	3.3	2.17	P; encodes Rab-like GTP-binding protein PARF, which interacts with ARF (or CDKN2A). Other Rab has been implicated in ID ⁴⁹ .
8500031	S	<i>CACNA1G</i>	S1346fs	2.7	18.76	P, E; encodes a low-voltage-activated calcium channel which may also modulate the firing patterns of neurons ^{23,24} .
8600057	S	<i>CAPN10</i>	138ins5aa	2.1	2.09	E; calcium-regulated non-lysosomal endopeptidase with a role in cytoskeletal remodelling and signal transduction, involved in long-term potentiation ⁵¹ .
8600495	NS	<i>CASP2</i>	Q392X	2.5	29.62	P; caspase 2, role in apoptosis, abnormal in <i>CASP2</i> -deficient mice, particularly for motor and sympathetic neurons ⁵² . Motor abnormalities not observed in family.
M346	NS	<i>CCNA2</i>	Splice site*	3.3	52.17	S, P; cyclin A2 is essential for cell cycle control ⁵³ . In mice, targeted deletion of this gene is lethal ⁵⁴ . Regulated by <i>SCAPER</i> , mutated in family 8600277.
8500235‡	S	<i>CNKSR1</i>	T282fs	2.53	15.83	P; regulates Raf in the MAPK pathway, acts as scaffold protein linking Ras and Rho signal transduction pathways ²⁹ . Interacts with RALGDS, which is mutated in family 8500155.
M144	NS	<i>COQ5</i>	G118S	1.8	15.10	P, E; methyltransferase with pivotal role in coenzyme Q biosynthesis. Interacts with NAB2 which controls length of poly(A) tail (see http://thebiogrid.org/35094/summary/saccharomyces-cerevisiae/coq5.html). The human orthologue of NAB2 is implicated in ARID ³² .
M178	NS	<i>EEF1B2</i>	Splice site*	2.6	13.84	S, P, E; controls translation by transferring aminoacyl-tRNAs to the ribosome. Interacts with UNC51-like kinase 2 which is involved in axonal elongation translation ⁵⁵ .
G017	NS	<i>ELP2</i>	T555P	2.4	14.33	P, E; encodes subunit of the RNA polymerase II elongator complex ⁵⁶ . ELP3 subunit implicated in motor neuron degeneration. Allelic <i>ELP2</i> mutation found in family M8500061.
8500061 M263	NS NS	<i>ELP2</i> <i>ENTPD1</i>	R462L Y65C	2.7 2.65	16.98 12.12	P, E; involved in transcriptional elongation, see also family G017 with allelic <i>ELP2</i> mutation. P, E; ectonucleoside triphosphate diphosphohydrolase, expressed in CNS; knockout mice display abnormal synaptic transmitter release ⁵⁷ .
M050†	S	<i>ERLIN2</i>	R36K	3.73	12.72	S, P, E; involved in the ER-associated degradation of inositol 1,4,5-triphosphate receptors ⁵⁸ .
8500058	NS	<i>FASN</i>	R1819W	3.3	4.50	P; gene product synthesizes long-chain fatty acids from acetyl-CoA and malonyl-CoA. Expressed in post-synaptic density. In mice, <i>FASN</i> deficiency leads to embryonic lethality ⁵⁹ .
M269	S	<i>FRY</i>	R1197X	2.8	12.68	P; regulates actin cytoskeleton, limits dendritic branching. In HeLa cells, <i>FRY</i> binds to microtubules and localizes on the spindle and is crucial for the alignment of mitotic chromosomes ⁶⁰ .
M251	S	<i>GON4L</i>	Splice site*	3.01	40.19	P, E; cloned from brain. Encodes a transcription factor thought to function in cell cycle control ⁶¹ .
M189‡	S	<i>HIST1H4B</i>	K9fs	2.1	48.87	P, E; encodes a member of the histone H4 family; analogy to histone H3 mutation in family G002. Ehlers–Danlos-related symptoms are probably due to <i>TNXB</i> mutation.
G002	NS	<i>HIST3H3</i>	R130C	2.53	26.74	P; role in spindle assembly and chromosome bi-orientation ^{62–64} . See also family M189 with <i>HIST1H4B</i> mutation.
8500064	NS	<i>INPP4A</i>	D915fs	2.4	46.16	P, E; encodes inositol polyphosphate-4-phosphatase, only plausible change in family. Regulates localization of synaptic NMDA receptors, protects neurons from excitotoxic cell death ⁶⁵ . Knockout mice develop locomotor instability; not observed in this family.
M061	S	<i>KDM5A</i>	R719G	2.3	6.06	P, E; encodes histone demethylase specific for Lys 4 of histone H3, role in transcriptional regulation ⁶⁶ . Other histone demethylase has been implicated in X-linked ID ¹⁶ . See also family M8303971 with <i>KDM6B</i> mutation.

in humans, adenosine kinase deficiency should lead to intellectual disability, whereas in the mouse, overexpression of *Adk* causes neurological symptoms, and *Adk* deficiency gives rise to early lethal liver steatosis³⁰. Nothing is known yet about the function of the *C12orf57* gene, apart from its apparent overlap with *ATN1* (see UCSC Genome Browser, NCBI36/hg18). CAG trinucleotide expansion in the *ATN1* gene is the cause of dentatorubral pallidoluysian atrophy (DRPLA), another syndromic form of intellectual disability. A comprehensive list of families with single, probably disease-causing mutations is shown in Table 2.

Despite exhaustive validation of our data and stringent filtering against all known neutral and pathogenic sequence variants (see

Supplementary Information and Supplementary Tables 3–6), it is still possible that not all of these changes will turn out to be causative. Particularly for the numerous missense mutations observed, functional studies will be required to rule out rare polymorphisms that are unrelated to intellectual disability. In a previous study, 1% of the protein-truncating mutations on the X chromosome were found to be unrelated to disease³¹, and in our study, 12 observed inactivating mutations did not co-segregate with intellectual disability (see Supplementary Table 4). However, we believe that the vast majority of the changes presented here as probably pathogenic will be confirmed, even if they have been observed only once, because most of the proteins encoded by these novel candidate genes interact with the

Table 2 | Continued

Family	Phenotype	Gene	Mutation	LOD score	Length (Mb)	Supporting evidence
8303971	S	<i>KDM6B</i>	P888S	3.1	5.08	S, P; demethylase 6B specifically targeting Lys 27 of histone H3, has a central role in regulation of posterior development by regulating HOX gene expression ⁶⁷ . Mutation of <i>KDM5A</i> gives rise to ID (see family M061).
M154	S	<i>KIF7</i>	E758K	2.1	7.46	P, E; knockout mouse model with complex picture involving brain and other neurological abnormalities ⁶⁸ . Stickler-like clinical features in this family can be explained by co-existing <i>COL9A1</i> mutation.
M183	S	<i>LAMA1</i>	G1572fs	2.1	5.82	S, P; codes for subunit of laminin, role in attachment, migration and organization of cells during embryonic development. Required for normal retinal development in mice ⁶⁹ .
G030	S	<i>LARP7</i>	K276fs	1.93	8.94	S,P; encodes negative transcriptional regulator of polymerase II genes ⁷⁰ .
7903104	S	<i>LINS1</i>	H329fs	2.65	7.87	S, P; similar to <i>lin</i> , a <i>Drosophila</i> gene having important roles in the development of the epidermis and the hindgut. Link with ID unclear.
8600060†	NS	<i>MAN1B1</i>	R334C	3.13	2.49	P, E; encodes mannosidase that targets misfolded glycoproteins for degradation. <i>MAN1B1</i> frameshift mutation observed in another ARID family by Canadian group (J. Vincent, personal communication).
8600277	NS	<i>NDST1</i>	R709Q	2.1	10.18	S, P; only change in family. Encodes heparan <i>N</i> -deacetylase/ <i>N</i> -sulphotransferase, deficiency is lethal in mice due to respiratory distress ⁷¹ . No obvious link with ID.
M158	S	<i>PARP1</i>	L293F	1.8	16.76	P; poly(ADP-ribose) polymerase involved in histone 1 modification; role in memory stabilization in mice ⁷² .
M194	NS, ASD	<i>PECR</i>	L57V	2.5	11.27	P; brain-expressed peroxisomal <i>trans</i> -2-enoyl-CoA reductase involved in the biosynthesis of unsaturated fatty acids ⁷³ .
8401214	S	<i>POLR3B</i>	T199K	1.93	24.89	E; second-largest core component of RNA polymerase III, which synthesizes small RNAs such as tRNAs and 5S rRNAs ²¹ .
8500302	NS	<i>PRMT10</i>	G189R	2.65	9.75	P, E; protein arginine methyltransferase 10. Protein arginine methylation affects chromatin remodelling leading to transcriptional regulation, RNA processing, DNA repair and cell signalling ⁷⁴ .
M010	NS	<i>PRRT2</i>	A214fs	5.2	25.59	P; interacts with SNAP25 which in turn assembles with syntaxin-1 and synaptobrevin to form exocytotic fusion complex in neurons ⁵⁵ .
8500155	NS	<i>RALGDS</i>	A706V	4.0	5.56	S, E; effector of Ras-related RalA and RalB GTPases, role in synaptic plasticity ²⁶ . Interacts with CNKSR1, inactivated in family 8500235.
8700136	NS, ASD	<i>RGS7</i>	N304fs	2.53	24.34	P; regulator of G protein signalling. Interacts with 14-3-3 protein, tau and snapin, a component of the SNARE complex required for synaptic vesicle docking and fusion ⁷⁵ . Indirectly linked with <i>ADRA2B</i> , mutated in family M266_2.
8600086	NS	<i>SCAPER</i>	Y118fs	3.9	17.45	S, E; interacts with CCNA2/CDK2 complex, transiently maintains CCNA2 in cytoplasm ⁷⁶ . CCNA2 is mutated in family M346.
8600012	S	<i>SLC31A1</i>	R90G	2.1	13.85	P, E; encodes one of two genes involved in copper import. Deficiency of the SLC31A1 orthologue in mice is early lethal, heterozygotes have progressive neurological disorder ⁷⁷ , similar to patients in this family.
M177	S	<i>TAF2</i>	W649R	2.1	19.16	P, E; TATA-box-associated gene is very important regulator of transcription (see OMIM 604912). Other TAF genes have been implicated in X-linked ID (V.K. <i>et al.</i> , manuscript in preparation). MAL2 is another, less likely, candidate in this family.
M160	S	<i>TMEM135</i>	C228S	2.4	16.89	S, P, E; transmembrane protein involved in fat metabolism and energy expenditure ⁷⁸ .
M300	NS	<i>TRMT1</i>	I230fs	3.4	10.34	P, E; encodes dimethylguanosine tRNA methyltransferase ⁷⁹ . At least two other RNA methyltransferases have been implicated in ID (ref. 17 and L.A.M., manuscript in preparation).
M168	NS, ASD	<i>UBR7</i>	N124S	2.5	8.78	P, E; encodes n-regognin 7, a component of E3 ubiquitin ligase ⁸⁰ . Involved in protein degradation, which has been implicated in ID.
8500320	S	<i>WDR45L</i>	R109Q	1.93	2.55	P, E; WD repeat domain, phosphoinositide-interacting protein 3, ILF1-like ⁸¹ , specific function unknown.
M169	S	<i>ZBTB40</i>	Q525X	3.5	14.56	S, P, E; krüppel-type zinc finger, highly expressed in brain. Regulator of glia differentiation ²⁵ .
M156	NS	<i>ZCCHC8</i>	L90X	2.3	7.64	P; zinc-finger protein, identified in the spliceosome C complex. Interacts with BRCA1 and RBM7 ^{82,83} . RBM10 has been implicated in X-linked ID (V.K. <i>et al.</i> , manuscript in preparation).
M025	NS	<i>ZNF526</i>	R459Q	4.5	6.13	P; zinc-finger protein, only remaining change in family. Functional relevance supported by 3D modelling. Probable activator of mRNA translation. Allelic <i>ZNF526</i> mutation observed in family 8500156.
8500156	NS	<i>ZNF526</i>	Q539H	4.04	11.33	P; see family M025 with allelic <i>ZNF526</i> mutation.

References 44–83 are listed in Supplementary Information. E, high evolutionary conservation score; P, high pathogenicity score, includes truncating mutations; S, only change found in family. ASD, autism spectrum disorder; GPCR, G-protein-coupled receptor; ID, intellectual disability; NS, non-syndromic; S, syndromic.

* See Supplementary Information for further details.

† Parents are distantly related. LOD scores provided are minimum estimates, calculated on the assumption that they are second cousins.

‡ In ethnically matching healthy controls a single heterozygous carrier was found (for details, see Supplementary Table 3).

products of known or novel genes associated with intellectual disability, as shown in Fig. 1.

Most ARID genes are not synapse specific

We have previously shown that ARID is an extremely heterogeneous disorder⁶. In contrast to non-syndromic hearing impairment or X-linked intellectual disability, common forms of ARID do not seem to exist, although there is evidence for regional clustering of the underlying gene defects⁵. Extrapolating from the number of known X-chromosomal intellectual disability genes argues for the involvement of several hundred genes in non-syndromic ARID, and the total number of ARID genes may well run into the thousands¹.

Identification of most or all of these genes is a prerequisite for early diagnosis, prevention and, eventually, therapy of intellectual disability, but at the present pace, many years would be required to accomplish this task. Here, we have combined homozygosity mapping, targeted exon enrichment and next-generation sequencing to speed up the molecular elucidation of ARID. In 78 out of 136 consanguineous families investigated, we have found apparently pathogenic mutations in single genes. Fifty of these genes had not been implicated in ARID before, and only two of these novel intellectual disability genes were found to be mutated in two independent families. None of the ~10 previously known genes for non-syndromic ARID, including those that were identified in Iranian families^{32–36}, was observed in our present

cohort, thereby corroborating previous evidence that ARID is extremely heterogeneous.

Much of the research into the molecular causes of intellectual disability has focused on the synapse and synapse-specific genes (for example, see refs 2, 37). In the present study, relatively few of the novel defects identified involve synapse- or neuron-specific genes, and they are vastly outnumbered by ubiquitously expressed genes with indispensable cellular functions, such as DNA transcription and translation, protein degradation, mRNA splicing, energy metabolism as well as fatty-acid synthesis and turnover. Many of these defects were found to be associated with non-syndromic ARID. It is not immediately clear why the clinical consequences of defects involving such a wide spectrum of basic cellular processes should be confined to the brain, but this conceivably reflects the complexity of the central nervous system which may render it particularly vulnerable to damage.

We expect that these findings will have direct implications for the diagnosis and prevention of intellectual disability, and perhaps also for autism, schizophrenia and epilepsy, which often co-exist in intellectual disability patients and are frequently associated with mutations in the same genes (for example, see ref. 38; reviewed in ref. 1). Further investigation of the novel genes and networks presented here should significantly deepen our insight into the pathogenesis of intellectual disability and related disorders. Moreover, this study illustrates the power of large-scale next-generation sequencing in families as a general strategy to shed light on the aetiology of complex disorders and on the function of the underlying genes.

Note added in proof: While this work was in the press, two unrelated groups reported on inactivating ERLIN2 mutations in patients with recessive intellectual disability and progressive motor dysfunction^{39,40}. Moreover, syndromic forms of intellectual disability have been described in patients with AP4B1 and AP4E1 (ref. 41) and MAN1B1 (ref. 42) mutations, respectively. Finally, mutations inactivating the *KIF7* gene were identified as the cause of the recessive fetal hydrocephalus and acrocallosal syndromes that include brain malformations⁴³.

METHODS SUMMARY

Most families studied were from Iran, and less than 10% had a Turkish or Arabic background. Wechsler Intelligence Scales for Children (WISC) and WAIS were used to assess the IQ in children and parents. Many of the pedigrees, as well as the methods used for autozygosity mapping, have been described previously.

Exons from homozygous intervals were enriched with custom-made Agilent SureSelect DNA capture arrays and sequenced on an Illumina Genome Analyser II yielding 76-bp single reads. >98% of the targeted exons were covered by at least four non-redundant sequence reads, each with a PHRED-like quality score of 20 or above (mean, 0.984; median, 0.993; for details, see Supplementary Table 5).

To assess the reliability of this procedure for calling homozygous mutations, we looked up SNP markers from homozygous intervals of five selected families that had been analysed with high-resolution SNP arrays. For 773 out of 776 markers, next-generation sequencing and array-based SNP typing yielded identical results.

To detect single nucleotide variants, high-quality reads were aligned to the human reference genome (hg18) by SOAP2.20 with default settings, typically gap-free. Homozygous exon-spanning deletions were assumed if the sequence coverage of the relevant exon(s) was reduced to <5% of the mean. Details about the detection of smaller deletions and insertions are provided in Methods. All variants were validated by high-resolution array CGH, Sanger sequencing, or both.

Homozygous variants were filtered against dbSNP130/131, whole genomes from 185 healthy individuals studied by the 1000 Genomes Project and exomes from 200 Danish individuals, and found to be absent in at least 100 chromosomes from Iranian controls (see Supplementary Tables 1 and 3). To select and prioritize apparently disease-causing variants, various criteria were used (for more details, see Methods). All putative mutations co-segregated with intellectual disability in the respective families.

Received 9 March; accepted 5 August 2011.

Published online 21 September 2011.

1. Ropers, H. H. Genetics of early onset cognitive impairment. *Annu. Rev. Genomics Hum. Genet.* **11**, 161–187 (2010).

2. Hamdan, F. F. *et al.* Excess of *de novo* deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **88**, 306–316 (2011).
3. Hamdan, F. F. *et al.* Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation. *N. Engl. J. Med.* **360**, 599–605 (2009).
4. Vissers, L. E. *et al.* A *de novo* paradigm for mental retardation. *Nature Genet.* **42**, 1109–1112 (2010).
5. Najmabadi, H. *et al.* Homozygosity mapping in consanguineous families reveals extreme heterogeneity of non-syndromic autosomal recessive mental retardation and identifies 8 novel gene loci. *Hum. Genet.* **121**, 43–48 (2007).
6. Kuss, A. W. *et al.* Autosomal recessive mental retardation: homozygosity mapping identifies 27 single linkage intervals, at least 14 novel loci and several mutation hotspots. *Hum. Genet.* **129**, 141–148 (2011).
7. Kaufman, L., Ayub, M. & Vincent, J. B. The genetic basis of non-syndromic intellectual disability: a review. *J. Neurodev. Disord.* **2**, 182–209 (2010).
8. Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature Genet.* **42**, 790–793 (2010).
9. Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nature Genet.* **42**, 30–35 (2010).
10. Lupski, J. R. *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N. Engl. J. Med.* **362**, 1181–1191 (2010).
11. Cantagrel, V. *et al.* SRD5A3 is required for converting polyprenol to dolichol and is mutated in a congenital glycosylation disorder. *Cell* **142**, 203–217 (2010).
12. Kahrizi, K. *et al.* Next generation sequencing in a family with autosomal recessive Kahrizi syndrome (OMIM 612713) reveals a homozygous frameshift mutation in SRD5A3. *Eur. J. Hum. Genet.* **19**, 115–117 (2011).
13. Raas-Rothschild, A. *et al.* A PEX6-defective peroxisomal biogenesis disorder with severe phenotype in an infant, versus mild phenotype resembling Usher syndrome in the affected parents. *Am. J. Hum. Genet.* **70**, 1062–1068 (2002).
14. Steinfeld, R. *et al.* Folate receptor alpha defect causes cerebral folate transport deficiency: a treatable neurodegenerative disorder associated with disturbed myelin metabolism. *Am. J. Hum. Genet.* **85**, 354–363 (2009).
15. Shoichet, S. A. *et al.* Mutations in the ZNF41 gene are associated with cognitive deficits: identification of a new candidate for X-linked mental retardation. *Am. J. Hum. Genet.* **73**, 1341–1354 (2003).
16. Jensen, L. R. *et al.* Mutations in the *JARID1C* gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am. J. Hum. Genet.* **76**, 227–236 (2005).
17. Freude, K. *et al.* Mutations in the *FTSJ1* gene coding for a novel S-adenosylmethionine-binding protein cause nonsyndromic X-linked mental retardation. *Am. J. Hum. Genet.* **75**, 305–309 (2004).
18. Ramser, J. *et al.* A splice site mutation in the methyltransferase gene *FTSJ1* in Xp11.23 is associated with non-syndromic mental retardation in a large Belgian family (MRX9). *J. Med. Genet.* **41**, 679–683 (2004).
19. Klein, U., Ramirez, M. T., Kobilka, B. K. & von Zastrow, M. A novel interaction between adrenergic receptors and the α -subunit of eukaryotic initiation factor 2B. *J. Biol. Chem.* **272**, 19099–19102 (1997).
20. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods* **7**, 575–576 (2010).
21. Geiduschek, E. P. & Kassavetis, G. A. The RNA polymerase III transcription apparatus. *J. Mol. Biol.* **310**, 1–26 (2001).
22. Hurov, K. E., Cotta-Ramusino, C. & Elledge, S. J. A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability. *Genes Dev.* **24**, 1939–1950 (2010).
23. Zamponi, G. W., Lory, P. & Perez-Reyes, E. Role of voltage-gated calcium channels in epilepsy. *Pflügers Arch.* **460**, 395–403 (2010).
24. Singh, B. *et al.* Mutational analysis of CACNA1G in idiopathic generalized epilepsy. *Hum. Mutat.* **28**, 524–525 (2007).
25. Södersten, E., Lilja, T. & Hermanson, O. The novel BTB/POZ and zinc finger factor Zbtb45 is essential for proper glial differentiation of neural and oligodendrocyte progenitor cells. *Cell Cycle* **9**, 4866–4875 (2010).
26. Zhu, X. *et al.* Disruption of PC1/3 expression in mice causes dwarfism and multiple neuroendocrine peptide processing defects. *Proc. Natl Acad. Sci. USA* **99**, 10293–10298 (2002).
27. Billuart, P. *et al.* Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation. *Nature* **392**, 923–926 (1998).
28. D'Adamo, P. *et al.* Mutations in GDI1 are responsible for X-linked non-specific mental retardation. *Nature Genet.* **19**, 134–139 (1998).
29. Jaffe, A. B., Aspenstrom, P. & Hall, A. Human CNK1 acts as a scaffold protein, linking Rho and Ras signal transduction pathways. *Mol. Cell. Biol.* **24**, 1736–1746 (2004).
30. Boison, D. *et al.* Neonatal hepatic steatosis by disruption of the adenosine kinase gene. *Proc. Natl Acad. Sci. USA* **99**, 6985–6990 (2002).
31. Tarpey, P. S. *et al.* A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nature Genet.* **41**, 535–543 (2009).
32. Pak, C. *et al.* Mutation of the conserved polyadenosine RNA binding protein, ZC3H14/dNab2, impairs neural function in *Drosophila* and humans. *Proc. Natl Acad. Sci. USA* **108**, 12390–12395 (2011).
33. Garshasbi, M. *et al.* A defect in the *TUSC3* gene is associated with autosomal recessive mental retardation. *Am. J. Hum. Genet.* **82**, 1158–1164 (2008).
34. Mir, A. *et al.* Identification of mutations in *TRAPPC9*, which encodes the NIK- and IKK- β -binding protein, in nonsyndromic autosomal-recessive mental retardation. *Am. J. Hum. Genet.* **85**, 909–915 (2009).
35. Moheb, L. A. *et al.* Identification of a nonsense mutation in the very low-density lipoprotein receptor gene (*VLDLR*) in an Iranian family with dysequilibrium syndrome. *Eur. J. Hum. Genet.* **16**, 270–273 (2008).

36. Motazacker, M. M. *et al.* A defect in the ionotropic glutamate receptor 6 gene (*GRIK2*) is associated with autosomal recessive mental retardation. *Am. J. Hum. Genet.* **81**, 792–798 (2007).
37. Laumonnier, F., Cuthbert, P. C. & Grant, S. G. The role of neuronal complexes in human X-linked brain diseases. *Am. J. Hum. Genet.* **80**, 205–220 (2007).
38. Ullmann, R. *et al.* Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum. Mutat.* **28**, 674–682 (2007).
39. Yildirim, Y. *et al.* A frameshift mutation of *ERLIN2* in recessive intellectual disability, motor dysfunction and multiple joint contractures. *Hum. Mol. Genet.* **20**, 1886–1892 (2011).
40. Alazami, A. M., Adly, N., Al Dhalaan, H. & Alkuraya, F. S. A nullimorphic *ERLIN2* mutation defines a complicated hereditary spastic paraplegia locus (SPG18). *Neurogenetics* doi:10.1007/s10048-011-0291-8 (2011).
41. Abou Jamra, R. *et al.* Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *Am. J. Hum. Genet.* **88**, 788–795 (2011).
42. Rafiq, M. A. *et al.* Mutations in the alpha 1,2-mannosidase gene, *MAN1B1*, cause autosomal-recessive intellectual disability. *Am. J. Hum. Genet.* **89**, 176–182 (2011).
43. Putoux, A. *et al.* KIF7 mutations cause fetal hydrolethals and acrocallosal syndromes. *Nature Genet.* **43**, 601–606 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We express our gratitude to the patients and their families for their participation in the study. We thank S. Nakhee and K. Javan for their support;

S. Arzhang, S. Banihashemi, M. Kasiri, H. Khodae, M. Schlicht and M. Gerloff for contributing to this project in various ways; G. Eder for her assistance with the preparation of the manuscript; and S. Shoichet for critically reading the manuscript. This work was supported by the Max Planck Innovation Funds, the German Federal Ministry of Education and Research through the MRNET (grant 01GS08161, to H.H.R.), the Iranian National Science Foundation and the EU-FP7 project GENCODYS.

Author Contributions H.H.R. and H.N. initiated and directed this study. H.H., M.G., W.C., S.H., K.W., V.K., R.U., K.K. and A.W.K. contributed to its design and coordination. H.N., K.K., A.T., P.J., V.H., D.W., M.C., I.R., F.M., C.H., A.D., A.R., M.J.S.B., M.F. and H.D. recruited patients and families, and K.K. and A.T. were responsible for the clinical investigations. F.B., S.G.F. and R.K. did the karyotyping. M.G. performed the linkage analyses and together with R.W. and H.H., he established data management tools. R.U. and I.M. performed exon enrichments, and C.J. and M.B. did the deep sequencing experiments. H.H. analysed the sequencing data and provided bioinformatics support. M.G., S.S.A., M.H., A.Z., M.M., L.P., L.N.V., L.A.M., F.L., B.L., S.E.-N., Z.F., J.H., L.M. and A.W.K. participated in the validation of the results and the segregation analyses. T.Z. performed the pathway analyses. H.H., M.G., A.Z., L.P., R.W., T.Z., L.M., A.W.K., A.T., K.K., H.N. and H.H.R. evaluated and interpreted the results, and H.H.R., H.H. and M.G. wrote the manuscript.

Author Information Raw sequencing data can be retrieved from the Sequence Read Archive (SRA), accession number SRA036250. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.H.R. (ropers@molgen.mpg.de) or K.K. (kkahrizi@uswr.ac.ir).

Frequent pathway mutations of splicing machinery in myelodysplasia

Kenichi Yoshida^{1*}, Masashi Sanada^{1*}, Yuichi Shiraishi^{2*}, Daniel Nowak^{3*}, Yasunobu Nagata^{1*}, Ryo Yamamoto⁴, Yusuke Sato¹, Aiko Sato-Otsubo¹, Ayana Kon¹, Masao Nagasaki⁵, George Chalkidis⁶, Yutaka Suzuki⁷, Masashi Shiosaka¹, Ryoichiro Kawahata¹, Tomoyuki Yamaguchi⁸, Makoto Otsu⁴, Naoshi Obara⁹, Mamiko Sakata-Yanagimoto⁹, Ken Ishiyama¹⁰, Hiraku Mori¹¹, Florian Nolte³, Wolf-Karsten Hofmann³, Shuichi Miyawaki¹⁰, Sumio Sugano⁷, Claudia Haerlach¹², H. Phillip Koeffler^{13,14}, Lee-Yung Shih¹⁵, Torsten Haerlach¹², Shigeru Chiba⁹, Hiromitsu Nakauchi^{4,8}, Satoru Miyano^{2,6} & Seishi Ogawa¹

Myelodysplastic syndromes and related disorders (myelodysplasia) are a heterogeneous group of myeloid neoplasms showing deregulated blood cell production with evidence of myeloid dysplasia and a predisposition to acute myeloid leukaemia, whose pathogenesis is only incompletely understood. Here we report whole-exome sequencing of 29 myelodysplasia specimens, which unexpectedly revealed novel pathway mutations involving multiple components of the RNA splicing machinery, including *U2AF35*, *ZRSR2*, *SRSF2* and *SF3B1*. In a large series analysis, these splicing pathway mutations were frequent (~45 to ~85%) in, and highly specific to, myeloid neoplasms showing features of myelodysplasia. Conspicuously, most of the mutations, which occurred in a mutually exclusive manner, affected genes involved in the 3'-splice site recognition during pre-mRNA processing, inducing abnormal RNA splicing and compromised haematopoiesis. Our results provide the first evidence indicating that genetic alterations of the major splicing components could be involved in human pathogenesis, also implicating a novel therapeutic possibility for myelodysplasia.

Myelodysplastic syndromes (MDS) and related disorders (myelodysplasia) comprise a group of myeloid neoplasms characterized by deregulated, dysplastic blood cell production and a predisposition to acute myeloid leukaemia (AML)¹. Although the prevalence of MDS has not been determined precisely, more than 10,000 people are estimated to develop myelodysplasia annually in the United States². Their indolent clinical course before leukaemic transformation and ineffective haematopoiesis with evidence of myeloid dysplasia indicate a pathogenesis distinct from that involved in *de novo* AML. Currently, a number of gene mutations and cytogenetic changes have been implicated in the pathogenesis of MDS, including mutations of *RAS*, *TP53* and *RUNX1*, and more recently *ASXL1*, *c-CBL*, *DNMT3A*, *IDH1/2*, *TET2* and *EZH2* (ref. 3). Nevertheless, mutations of this set of genes do not fully explain the pathogenesis of MDS because they are also commonly found in other myeloid malignancies and roughly 20% of MDS cases have no known genetic changes (ref. 4 and unpublished data). In particular, the genetic alterations responsible for the dysplastic phenotypes and ineffective haematopoiesis of myelodysplasia are poorly understood. Meanwhile, the recent development of massively parallel sequencing technologies has provided an expanded opportunity to discover genetic changes across the entire genomes or protein-coding sequences in human cancers at a single-nucleotide level^{5–10}, which could be successfully applied to the genetic analysis of myelodysplasia to obtain a better understanding of its pathogenesis.

Overview of genetic alterations

In this study, we performed whole-exome sequencing of paired tumour/control DNA from 29 patients with myelodysplasia (Supplementary Table 1). Although incapable of detecting non-coding mutations and gene rearrangements, the whole-exome approach is a well-established strategy for obtaining comprehensive registries of protein-coding mutations at low cost and high performance. With a mean coverage of 133.8, 80.4% of the target sequences were analysed at more than $\times 20$ depth on average (Supplementary Fig. 1). All the candidates for somatic mutations ($N = 497$) generated through our data analysis pipeline were subjected to validation using Sanger sequencing (Supplementary Methods I and Supplementary Fig. 2). Finally, 268 non-synonymous somatic mutations were confirmed with an overall true positive rate of 53.9% (Supplementary Fig. 3), including 206 missense, 25 nonsense, and 10 splice site mutations, and 27 frameshift-causing insertions/deletions (indels) (Supplementary Fig. 4). The mutation rate of 9.2 (0–21) per sample was significantly lower than that in solid tumours (16.2–302)^{7,11,12} and multiple myeloma (32.4)⁶, but was comparable to that in AML (7.3–13)^{13–15} and chronic lymphocytic leukaemia (11.5)¹⁶. Combined with the genomic copy number profile obtained by single nucleotide polymorphism (SNP) array karyotyping, this array of somatic mutations provided a landscape of myelodysplasia genomes (Supplementary Fig. 5)^{17,18}.

¹Cancer Genomics Project, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655, Japan. ²Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ³Department of Hematology and Oncology, Medical Faculty Mannheim of the University of Heidelberg, 1-3 Theodor-Kutzer-Ufer, Mannheim 68167, Germany. ⁴Division of Stem Cell Therapy, Center for Stem Cell Biology and Regenerative Medicine, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁵Laboratory of Functional Genomics, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁶Laboratory of Sequence Data Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁷Division of Systems Biomedical Technology, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁸Nakauchi Stem Cell and Organ Regeneration Project, Exploratory Research for Advanced Technology, Japan Science and Technology Agency, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. ⁹Department of Hematology, Institute of Clinical Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba-shi, Ibaraki, 305-8571, Japan. ¹⁰Division of Hematology, Tokyo Metropolitan Ohtsuka Hospital, 2-8-1 Minami-Ohtsuka, Toshima-ku, Tokyo 170-0005, Japan. ¹¹Division of Hematology, Internal Medicine, Showa University Fujigaoka Hospital, 1-30 Fujigaoka, Aoba-ku, Yokohama, Kanagawa 227-8501, Japan. ¹²Munich Leukemia Laboratory, Max-Lebsche-Platz 31, Munich 81377, Germany. ¹³Hematology/Oncology, Cedars-Sinai Medical Center, 8700 Beverly Blvd, Los Angeles, California 90048, USA. ¹⁴National University of Singapore, Cancer Science Institute of Singapore, 28 Medical Drive, Singapore 117456, Singapore. ¹⁵Division of Hematology-Oncology, Department of Internal Medicine, Chang Gung Memorial Hospital, Chang Gung University, 199 Tung Hwa North Rd, Taipei 105, Taiwan.

*These authors contributed equally to this work.

Novel gene targets in myelodysplasia

The list of the somatic mutations (Supplementary Table 2) included most of the known gene targets in myelodysplasia with similar mutation frequencies to those previously reported, indicating an acceptable sensitivity of the current study. The mutations of the known gene targets, however, accounted for only 12.3% of all detected mutations ($N = 33$), and the remaining 235 mutations involved previously unreported genes. Among these, recurrently mutated genes in multiple cases are candidate targets of particular interest, for which high mutation rates are expected in general populations. In fact, 8 of the 12 recurrently mutated genes were among the well-described gene targets in myelodysplasia (Supplementary Table 3). However, what immediately drew our attention were the recurrent mutations involving *U2AF35* (also known as *U2AF1*), *ZRSR2* and *SRSF2* (SC35), because they belong to the common pathway known as RNA splicing. Including an additional three genes mutated in single cases (*SF3A1*, *SF3B1* and *PRPF40B*), six components of the splicing machinery were mutated in 16 out of the 29 cases (55.2%) in a mutually exclusive manner (Fig. 1, Supplementary Fig. 6 and Supplementary Table 2).

Frequent mutations in splicing machinery

RNA splicing is accomplished by a well-ordered recruitment, rearrangement and/or disengagement of a set of small nuclear ribonucleoprotein (snRNP) complexes (U1, U2, and either U4/5/6 or U11/12), as well as many other protein components onto the pre-mRNAs. Notably, the mutated components of the spliceosome were all engaged in the initial steps of RNA splicing, except for *PRPF40B*, whose functions in RNA splicing are poorly defined. Making physical interactions with SF1 and a serine/arginine-rich (SR) protein, such as *SRSF1* or *SRSF2*, the U2 auxiliary factor (U2AF) that consists of the U2AF65 (U2AF2)–U2AF35 heterodimer, is involved in the recognition of the 3' splice site (3'SS) and its nearby polypyrimidine tract, which is thought to be required for the subsequent recruitment of the U2 snRNP, containing *SF3A1* as well as *SF3B1*, to establish the splicing A complex (Fig. 1)¹⁹. *ZRSR2* (or *Urp*), is another essential component of the splicing machinery. Showing a close structural similarity to U2AF35, *ZRSR2* physically interacts with U2AF65, as well as *SRSF1* and *SRSF2*, with a distinct function from its homologue, U2AF35 (ref. 20).

To confirm and extend the initial findings in the whole-exome sequencing, we studied mutations of the above six genes together with

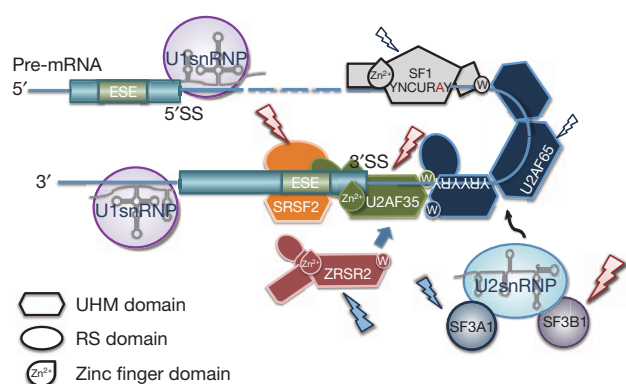


Figure 1 | Components of the splicing E/A complex mutated in myelodysplasia. RNA splicing is initiated by the recruitment of U1 snRNP to the 5'SS. SF1 and the larger subunit of the U2 auxiliary factor (U2AF), U2AF65, bind the branch point sequence (BPS) and its downstream polypyrimidine tract, respectively. The smaller subunit of U2AF (U2AF35) binds to the AG dinucleotide of the 3'SS, interacting with both U2AF65 and a SR protein, such as *SRSF2*, through its UHM and RS domain, comprising the earliest splicing complex (E complex). *ZRSR2* also interacts with U2AF and SR proteins to perform essential functions in RNA splicing. After the recognition of the 3'SS, U2 snRNP, together with *SF3A1* and *SF3B1*, is recruited to the 3'SS to generate the splicing complex A. The mutated components in myelodysplasia are indicated by arrows.

three additional spliceosome-related genes, including *U2AF65*, *SF1* and *SRSF1*, in a large series of myeloid neoplasms ($N = 582$) using a high-throughput mutation screen of pooled DNA followed by confirmation/identification of candidate mutations (refs 21 and 22 and Supplementary Methods II).

In total, 219 mutations were identified in 209 out of the 582 specimens of myeloid neoplasms through validating 313 provisional positive events in the pooled DNA screen (Supplementary Tables 4 and 5). The mutations among four genes, *U2AF35* ($N = 37$), *SRSF2* ($N = 56$), *ZRSR2* ($N = 23$) and *SF3B1* ($N = 79$), explained most of the mutations with much lower mutational rates for *SF3A1* ($N = 8$), *PRPF40B* ($N = 7$), *U2AF65* ($N = 4$) and *SF1* ($N = 5$) (Fig. 2). Mutations of the splicing machinery were highly specific to diseases showing myelodysplastic features, including MDS either with (84.9%) or without (43.9%) increased ring sideroblasts, chronic myelomonocytic leukaemia (CMML) (54.5%), and therapy-related AML or AML with myelodysplasia-related changes (25.8%), but were rare in *de novo* AML (6.6%) and myeloproliferative neoplasms (MPN) (9.4%) (Fig. 3a). The mutually exclusive pattern of the mutations in these splicing pathway genes was confirmed in this large case series, suggesting a common impact of these mutations on RNA splicing and the pathogenesis of myelodysplasia (Fig. 3b). The frequencies of mutations showed significant differences across disease types. Surprisingly, *SF3B1* mutations were found in the majority of the cases with MDS characterized by increased ring sideroblasts, that is, refractory anaemia with ring sideroblasts (RARS) (19/23 or 82.6%) and refractory cytopenia with multilineage dysplasia with $\geq 15\%$ ring sideroblasts (RCMD-RS) (38/50 or 76%) with much lower mutation frequencies in other myeloid neoplasms. RARS and RCMD-RS account

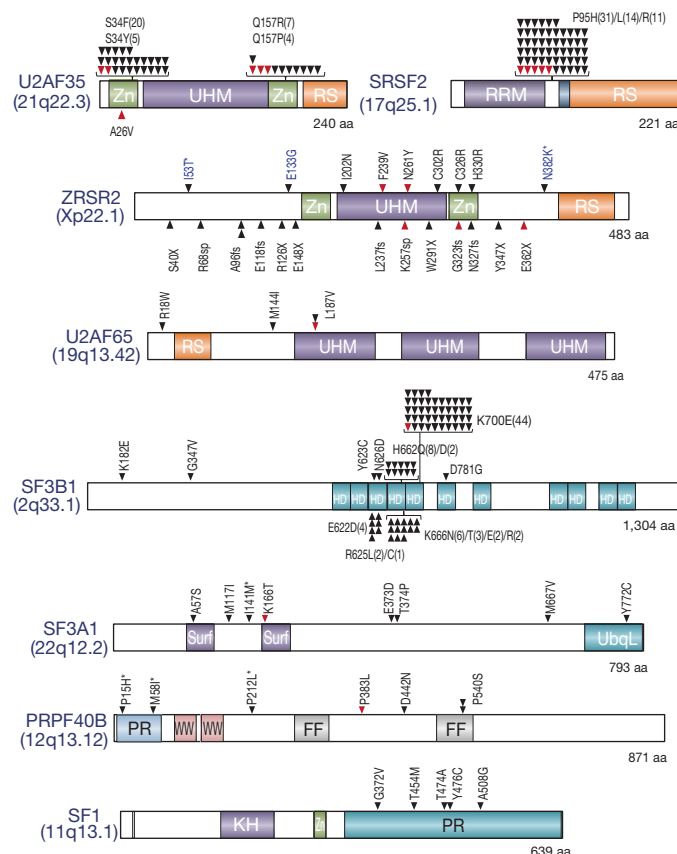
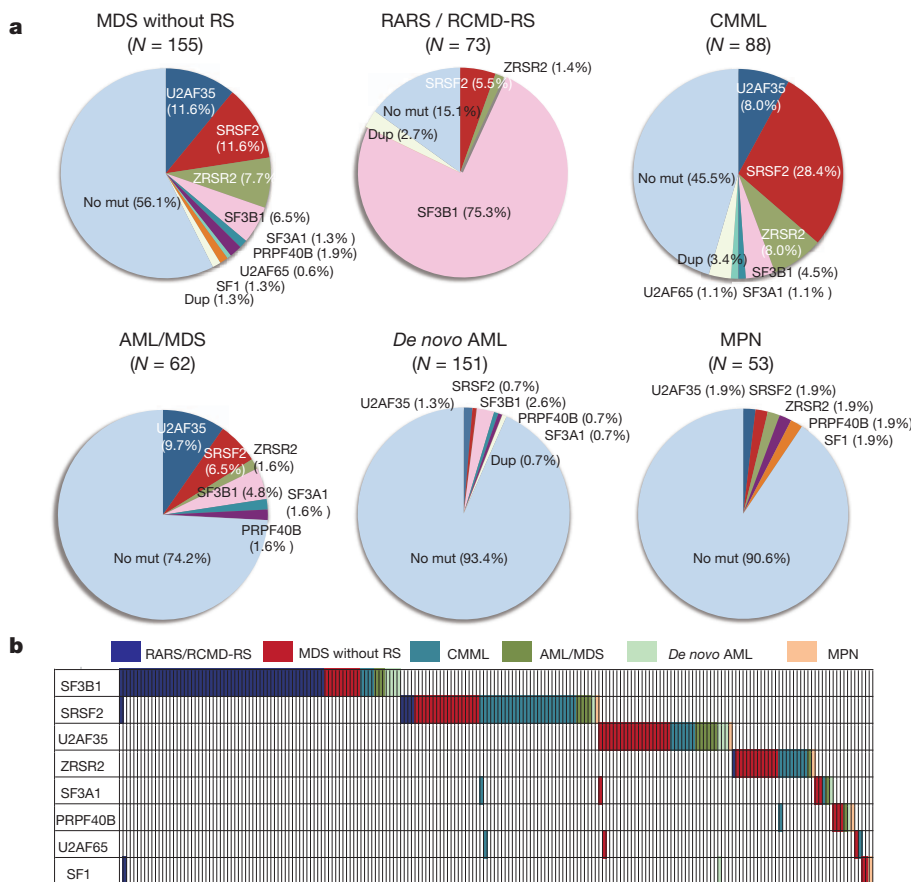


Figure 2 | Mutations of multiple components of the splicing machinery. Each mutation in the eight spliceosome components is shown with an arrowhead. Confirmed somatic mutations are discriminated by red arrows. Known domain structures are shown in coloured boxes as indicated. Mutations predicted as SNPs by MutationTaster (<http://www.mutationtaster.org/>) are indicated by asterisks. The number of each mutation is indicated in parenthesis. *ZRSR2* mutations in females are shown in blue.



for 4.3% and 12.9% of MDS cases, respectively, where deregulated iron metabolism has been implicated in the development of refractory anaemia²³. With such high mutation frequencies and specificity, the *SF3B1* mutations were thought to be almost pathognomonic to these MDS subtypes characterized by increased ring sideroblasts, and strongly implicated in the pathogenesis of MDS in these categories. Less conspicuously but significantly, *SRSF2* mutations were more frequent in CMML cases (Fig. 3 and Supplementary Table 4). Thus, although commonly involving the E/A splicing complexes, different mutations may still have different impacts on cell functions, contributing to the determination of discrete disease phenotypes. For example, studies have demonstrated that *SRSF2* was also involved in the regulation of DNA stability and that depletion of *SRSF2* can lead to genomic instability²⁴. Of interest in this context, regardless of disease subtypes, samples with *SRSF2* mutations were shown to have significantly more mutations of other genes compared with *U2AF35* mutations ($P = 0.001$, multiple regression analysis) (Supplementary Table 6 and Supplementary Fig. 7).

Notably, with a rare exception of A26V in a single case, the mutations of *U2AF35* exclusively involved two highly conserved amino acid positions (S34 or Q157) within the amino- and the carboxyl-terminal zinc finger motifs flanking the U2AF homology motif (UHM) domain. *SRSF2* mutations exclusively occurred at P95 within an intervening sequence between the RNA recognition motif (RRM) and arginine/serine-rich (RS) domains (Fig. 2 and Supplementary Figs 8 and 9). Similarly, *SF3B1* mutations predominantly involved K700 and, to a lesser extent, K666, H662 and E622, which are also conserved across species (Fig. 2 and Supplementary Fig. 10). The involvement of recurrent amino acid positions in these spliceosome genes strongly indicated a gain-of-function nature of these mutations, which has been a well-documented scenario in other oncogenic mutations²⁵. On the other hand, the 23 mutations in *ZRSR2* (Xp22.1) were widely distributed along the entire coding region (Fig. 2). Among these, 14 mutations were nonsense or frameshift changes, or involved splicing donor/acceptor

sites that caused either a premature truncation or a large structural change of the protein, leading to loss-of-function. Combined with their strong male preference for the mutation (14/14 cases), *ZRSR2* most likely acts as a tumour suppressor gene with an X-linked recessive mode of genetic action. The remaining nine *ZRSR2* mutations were missense changes and found in both males (six cases) and females (three cases), whose somatic origin was only confirmed in two cases. However, neither the dbSNP database (build131 and 132) nor the 1000 Genomes database (May 2011 snp calls) contained these missense nucleotides, suggesting that many, if not all, of these missense changes are likely to represent functional somatic changes, especially those found in males. Interrogation of these hot spots for mutations in *U2AF35* and *SRSF2* found no mutations among lymphoid neoplasms, including acute lymphoblastic leukaemia ($N = 24$) or non-Hodgkin's lymphoma ($N = 87$) (data not shown).

RNA splicing and spliceosome mutations

Because the splicing pathway mutations in myelodysplasia widely and specifically affect the major components of the splicing complexes E/A in a mutually exclusive manner, the common consequence of these mutations is logically the impaired recognition of 3'SSs that would lead to the production of aberrantly spliced mRNA species. To appreciate this and also to gain an insight into the biological/biochemical impact of these splicing mutations, we expressed the wild-type and the mutant (S34F) *U2AF35* in HeLa cells using retrovirus-mediated gene transfer with enhanced green fluorescent protein (EGFP) marking (Fig. 4a and Supplementary Methods III) and examined their effects on gene expression in these cells using GeneChip Human genome U133 plus 2.0 arrays (Affymetrix), followed by gene set enrichment analysis (GSEA) (Supplementary Methods IV)²⁶. Intriguingly, the GSEA disclosed a significant enrichment of the genes on the non-sense-mediated mRNA decay (NMD) pathway among the significantly upregulated genes in mutant *U2AF35*-transduced HeLa cells (Fig. 4b, Supplementary Fig. 11a and Supplementary Table 7), which was

confirmed by quantitative polymerase chain reactions (qPCR) (Fig. 4c and Supplementary Methods 5V). A similar result was also observed for the gene expression profile of an MDS-derived cell line (TF-1) transduced with the S34F mutant (Supplementary Figs 11b, c). The NMD activation by the mutant U2AF35 was suppressed significantly by the co-expression of the wild-type protein (Supplementary Fig. 11d), indicating that the effect of the mutant protein was likely to be mediated by inhibition of the functions of the wild-type protein. Given that the NMD pathway, known as mRNA surveillance, provides a post-transcriptional mechanism for recognizing and eliminating abnormal transcripts that prematurely terminate translation²⁷, the result of the GSEA analyses indicated that the mutant U2AF35 induced abnormal RNA splicing in HeLa and TF-1 cells, leading to the generation of unspliced RNA species having a premature stop codon and induction of the NMD activity.

To confirm this, we next performed whole transcriptome analysis in these cells using the GeneChip Human exon 1.0 ST Array (Affymetrix), in which we differentially tracked the behaviour of two discrete sets of probes showing different level of evidence of being exons, that is, 'Core' (authentic exons) and 'non-Core' (more likely introns) sets (Supplementary Methods IV and Supplementary Fig. 12). As shown in Fig. 4d, the Core and non-Core set probes were differentially enriched among probes showing significant difference in expression between wild-type and mutant-transduced cells (false discovery rate (FDR) = 0.01). The Core set probes were significantly enriched in those probes significantly downregulated in mutant U2AF35-transduced cells compared with wild-type U2AF35-transduced cells, whereas the non-Core set probes were enriched in those probes significantly upregulated in mutant U2AF35-transduced cells (Fig. 4e). The significant differential enrichment was also demonstrated, even when all probe sets were included (Fig. 4f). Moreover, the significantly differentially expressed Core set probes tended to be up- and downregulated in wild-type and mutant U2AF35-transduced cells compared with mock-transduced cells, respectively, and vice versa for the differentially expressed non-Core set probes (Fig. 4e). Combined, these exon array results indicated that the wild-type U2AF35 correctly promoted authentic RNA splicing, whereas the mutant U2AF35 inhibited this processes, rendering non-Core and therefore, more likely intronic sequences to remain unspliced.

The abnormal splicing in mutant U2AF35-transduced cells was more directly demonstrated by sequencing mRNAs extracted from HeLa cells, in which expression of the wild-type and mutant (S34F) U2AF35 were induced by doxycycline. First, after adjusting by the total number of mapped reads, the wild-type U2AF35-transduced cells showed an increased read counts in the exon fraction, but reduced counts in other fractions, compared with mutant U2AF35-transduced cells (Fig. 4g). The reads from the mutant-transduced cells were mapped to broader genomic regions compared with those from the wild-type U2AF35-transduced cells, which were largely explained by non-exon reads (Fig. 4h). Finally, the number of those reads that encompassed the authentic exon/intron junctions was significantly increased in mutant U2AF35-transduced cells compared with wild-type U2AF35-transduced cells (Fig. 4i and Supplementary Methods VI). These results clearly demonstrated that failure of splicing ubiquitously occurred in mutant U2AF35-transduced cells. A typical example of abnormal splicing in mutant-transduced cells and the list of significantly unspliced exons are shown in Supplementary Fig. 13 and Supplementary Table 8, respectively.

Biological consequence of U2AF35 mutations

Finally, we examined the biological effects of compromised functions of the E/A splicing complexes. First, TF-1 and HeLa cells were transduced with lentivirus constructs expressing either the S34F U2AF35 mutant or wild-type U2AF35 under a tetracycline-inducible promoter (Fig. 5a and Supplementary Figs 14a and 15a), and cell proliferation was examined after the induction of their expression. Unexpectedly, after the induction of gene expression with

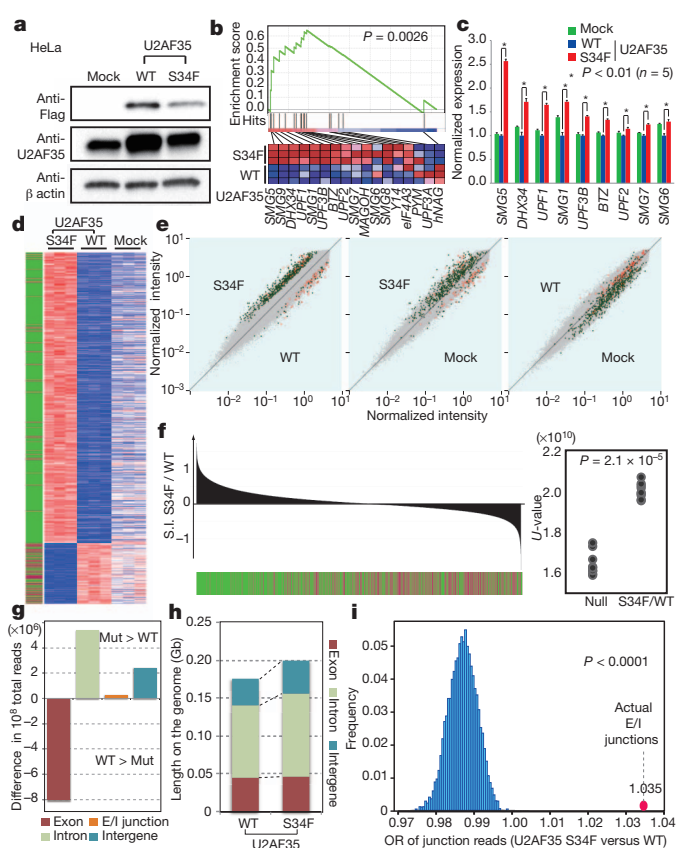


Figure 4 | Altered RNA splicing caused by a U2AF35 mutant. **a**, Western blot analyses showing expression of transduced wild-type or mutant (S34F) U2AF35 in HeLa cells used for the analyses of expression and exon microarrays. **b**, The GSEA demonstrating a significant enrichment of the set of 17 NMD pathway genes among significantly differentially expressed genes between wild-type and mutant U2AF35-transduced HeLa cells. The significance of the gene set was empirically determined by 1,000 gene-set permutations. **c**, The confirmation of the microarray analysis for the expression of nine genes that contributed to the core enrichment in the NMD gene set. Means \pm s.e. are provided for the indicated NMD genes. *P* values were determined by the Mann–Whitney *U* test. **d**, Significantly upregulated and downregulated probe sets (FDR = 0.01) in mutant U2AF35-transduced cells compared with wild-type U2AF35-transduced cells in triplicate exon array experiments are shown in a heat map. The origin of each probe set is depicted in the left lane, where red and green bars indicate the Core and non-Core sets, respectively. **e**, Pair-wise scatter plots of the normalized intensities of entire probe sets (grey) across different experiments. The Core and non-Core set probes that were significantly differentially expressed between the wild-type and mutant U2AF35-transduced cells are plotted in red and green, respectively. **f**, Distribution of the Core (red) and non-Core (green) probe sets within the entire probe sets ordered by splicing index (S.I.; Supplementary Methods IV), calculated between wild-type and mutant U2AF35-transduced cells. In the right panel, the differential enrichment of both probe sets was confirmed by Mann–Whitney *U* test. **g**, Difference in read counts for the indicated fractions per 10^8 total reads in RNA sequencing between wild-type and mutant U2AF35-expressing HeLa cells analysis. Increased/decreased read counts in mutant U2AF35-expressing cells are plotted upward/downward, respectively. **h**, Comparison of the genome coverage by the indicated fractions in wild-type- and mutant-U2AF35-expressing cells. The genome coverage was calculated for each fraction within the 10^8 reads randomly selected from the total reads and averaged for ten independent selections. **i**, The odds ratio of the junction reads within the total mapped reads was calculated between the two experiments (red circle), which was evaluated against the 10,000 simulated values under the null hypothesis (histogram in blue).

doxycycline, the mutant U2AF35-transduced cells, but not the wild-type U2AF35-transduced cells, showed reduced cell proliferation (Fig. 5b and Supplementary Fig. 15b) with a marked increase in the G2/M fraction (G2/M arrest) together with enhanced apoptosis as

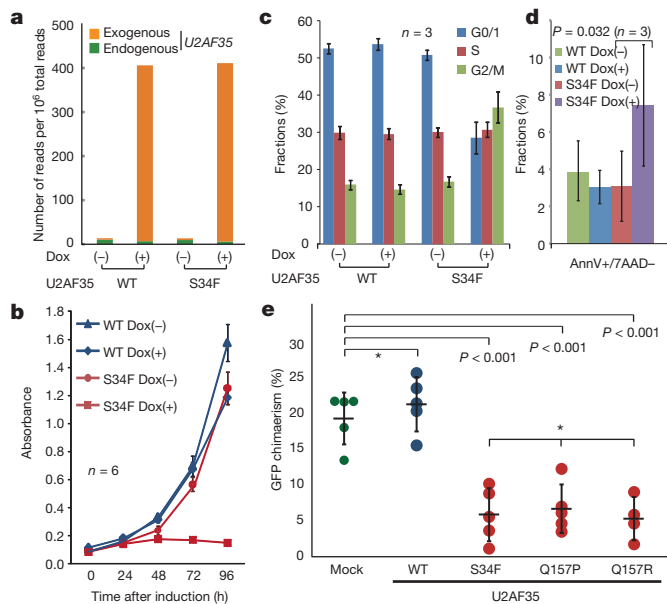


Figure 5 | Functional analysis of mutant U2AF35. **a**, Expression of endogenous and exogenous *U2AF35* transcripts in HeLa cells before and after induction determined by RNA sequencing. *U2AF35* transcripts were differentially enumerated for endogenous and exogenous species, which were discriminated by the Flag sequence. **b**, Cell proliferation assays of *U2AF35*-transduced HeLa cells, where cell numbers were measured using cell-counting apparatus and are plotted as mean absorbance \pm s.d. **c**, The flow cytometry analysis of propidium iodide (PI)-stained HeLa cells transduced with the different *U2AF35* constructs. Mean fractions \pm s.d. in G0/G1, S and G2/M populations after the induction of *U2AF35* expression are plotted. **d**, Fractions of the annexin V-positive (AnnV+) populations among the 7-amino-actinomycin D (7AAD)-negative population before and after the induction of *U2AF35* expression are plotted as mean \pm s.d. for indicated samples. The significance of difference was determined by paired *t*-test. **e**, Competitive reconstitution assays for CD34-negative KSL cells transduced with indicated *U2AF35* mutants. Chimaerism in the peripheral blood 6 weeks after transplantation are plotted as mean %EGFP-positive Ly5.1 cells \pm s.d., where outliers were excluded from the analysis. The significance of differences was evaluated by the Grubbs test with Bonferroni's correction for multiple testing. *not significant.

indicated by the increased sub-G1 fraction and annexin V-positive cells (Fig. 5c, d, Supplementary Fig. 14b and Supplementary Methods VI). To confirm the growth-suppressive effect of *U2AF35* mutants *in vitro*, a highly purified haematopoietic stem cell population (CD34⁺c-Kit⁺Scal⁺Lin⁻, CD34⁺KSL) prepared from C57BL/6 (B6)-Ly5.1 mouse bone marrow²⁸ was retrovirally transduced with either the mutant (S34F, Q157P and Q157R) or wild-type *U2AF35*, or the mock constructs, each harbouring the EGFP marker gene (Supplementary Fig. 16). The ability of these transduced cells to reconstitute the haematopoietic system was tested in a competitive reconstitution assay. The transduced cells were mixed with whole bone marrow cells from B6-Ly5.1/5.2 F1 mice, transplanted into lethally irradiated B6-Ly5.2 recipients, and peripheral blood chimaerism derived from EGFP-positive cells was assessed 6 weeks after transplantation by flow cytometry. We confirmed that each recipient mouse received comparable numbers of EGFP-positive cells among the different retrovirus groups by estimating the percentage of EGFP-positive cells and overall proliferation in transduced cells by *ex vivo* tracking. Also no significant difference was observed in their homing capacity to bone marrow as assessed by transwell migration assays (Supplementary Fig. 17). As shown in Fig. 5e, the wild-type *U2AF35*-transduced cells showed a slightly higher reconstitution capacity than the mock-transduced cells. On the other hand, the recipients of the cells transduced with the various *U2AF35* mutants showed significantly lower EGFP-positive cell chimaerism than those of either the mock- or the wild-type *U2AF35*-transduced

cells, indicating a compromised reconstitution capacity of the haematopoietic stem/progenitor cells expressing the *U2AF35* mutants. In summary, these mutants lead to loss-of-function of *U2AF35* most probably by acting in a dominant-negative fashion to the wild-type protein.

Discussion

Our whole-exome sequencing study unexpectedly unmasked a complexity of novel pathway mutations found in approximately 45% to 85% of myelodysplasia patients depending on the disease subtypes, which affected multiple but distinctive components of the splicing machinery and, as such, demonstrated the unquestionable power of massively parallel sequencing technologies in cancer research.

The RNA splicing system comprises essential cellular machinery, through which eukaryotes can achieve successful transcription and guarantee the functional diversity of their protein species using alternative splicing in the face of a limited number of genes²⁹. Accordingly, the meticulous regulation of this machinery should be indispensable for the maintenance of cellular homeostasis³⁰, deregulation of which causes severe developmental abnormalities^{31,32}. The current discovery of frequent mutations of the splicing pathway in myelodysplasia, therefore, represents another remarkable example that illustrates how cancer develops by targeting critical cellular functions. It also provides an intriguing insight into the mechanism of 'cancer specific' alternative splicing, which have long been implicated in the development of cancer, including MDS and other haematopoietic neoplasms^{33,34}.

In myelodysplasia, the major targets of spliceosome mutations seemed to be largely confined to the components of the E/A splicing complex, among others to *SF3B1*, *SRSF2*, *U2AF35* and *ZRSR2*, and to a lesser extent, to *SF3A1*, *SF1*, *U2AF65* and *PRPF40B*. The broad coverage of the wide spectrum of spliceosome components in our exome sequencing was likely to preclude frequent involvement of other components on this pathway (Supplementary Fig. 18). The surprising frequency and specificity of these mutations in this complex, together with the mutually exclusive manner they occurred, unequivocally indicate that the compromised function of the E/A complex is a hallmark of this unique category of myeloid neoplasms, playing a central role in the pathogenesis of myelodysplasia. The close relationship between the mutation types and unique disease subtypes also support their pivotal roles in MDS.

Given the critical functions of the E/A splicing complex on the precise 3'SS recognition, the logical consequence of these relevant mutations would be the impaired splicing involving diverse RNA species. In fact, when expressed in HeLa cells, the mutant *U2AF35* induced global abnormalities of RNA splicing, leading to increased production of transcripts having unspliced intronic sequences. On the other hand, the functional link between the abnormal splicing of RNA species and the phenotype of myelodysplasia is still unclear. Mutant *U2AF35* seemed to suppress cell growth/proliferation and induce apoptosis rather than confer a growth advantage or promote clonal selection. *ZRSR2* knockdown in HeLa cells has been reported to also result in reduced viability, arguing for the common consequence of these pathway mutations³⁵. These observations suggested that the oncogenic actions of these splicing pathway mutations are distinct from what is expected for classical oncogenes, such as mutated kinases and signal transducers, but could be more related to cell differentiation. Of note in this regard, the commonest clinical presentation of MDS is severe cytopenia in multiple cell lineages due to ineffective haematopoiesis with increased apoptosis rather than unlimited cell proliferation¹. In this regard, lessons may be learned from the recent findings on the pathogenesis of the 5q- syndrome, where haploinsufficiency of *RPS14* leads to increased apoptosis of erythroid progenitors, but not myeloproliferation^{36,37}.

A lot of issues remain to be answered, however, to establish the functional link between these splicing pathway mutations and the

pathogenesis of MDS, where the broad spectrum of RNA species affected by impaired splicing hampers identification of responsible gene targets. Moreover, the mutated components of the splicing machinery have distinct function of their own other than direct regulation of RNA splicing, involved in elongation and DNA stability, which may be important to determine specific disease phenotypes. Clearly, more studies are required to answer these questions through understanding of the molecular basis of their oncogenic actions.

METHODS SUMMARY

Whole-exome sequencing of paired tumour/normal DNA samples from the 29 patients was performed after informed consent was obtained. SNP array-based copy number analysis was performed as previously described^{17,18}. Mutation analysis of the splicing pathway genes in a set of 582 myeloid neoplasms were performed by first screening mutations in PCR-amplified pooled targets from 12 individuals, followed by validation/identification of the candidate mutations within the corresponding 12 individuals by Sanger sequencing. Flag-tagged cDNAs of the wild-type and mutant *U2AF35* were generated by *in vitro* mutagenesis, constructed into a murine stem cell virus-based retroviral vector as well as a tetracycline-inducible lentivirus-based expression vector, and used for gene transfer to CD34⁺KSL cells and cultured cell lines, with EGFP marking, respectively. Total RNA was extracted from wild-type or mutant *U2AF35*-transduced HeLa and TF-1 cells, and analysed on microarrays. RNA sequencing was performed according to the manufacturer's instructions (Illumina). Cell proliferation assays (MTT assays) on HeLa and TF-1 cells stably transduced with lentivirus *U2AF35* constructs were performed in the presence or absence of doxycycline. For competitive reconstitution assays, CD34⁺KSL cells collected from C57BL/6 (B6)-Ly5.1 mice were retrovirally transduced with various *U2AF35* constructs with EGFP marking, and transplanted with competitor cells (B6-Ly5.1/5.2 F1 mouse origin) into lethally irradiated B6-Ly5.2 mice 48 h after gene transduction. Frequency of EGFP-positive cells was assessed in peripheral blood by flow cytometry 6 weeks after the transplantation (Supplementary Methods VII). The primer sets used for validation of gene mutations and qPCR of NMD gene expression are listed in Supplementary Tables 9–11. A complete description of the materials and methods is provided in the Supplementary Information. This study was approved by the ethics boards of the University of Tokyo, Munich Leukaemia Laboratory, University Hospital Mannheim, University of Tsukuba, Tokyo Metropolitan Ohtsuka Hospital and Chang Gung Memorial Hospital. Animal experiments were performed with approval of the Animal Experiment Committee of the University of Tokyo.

Received 7 June; accepted 24 August 2011.

Published online 11 September 2011.

- Corey, S. J. *et al.* Myelodysplastic syndromes: the complexity of stem-cell diseases. *Nature Rev. Cancer* **7**, 118–129 (2007).
- Ma, X., Does, M., Raza, A. & Mayne, S. T. Myelodysplastic syndromes: incidence and survival in the United States. *Cancer* **109**, 1536–1542 (2007).
- Bejar, R., Levine, R. & Ebert, B. L. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J. Clin. Oncol.* **29**, 504–515 (2011).
- Sanada, M. *et al.* Gain-of-function of mutated *C-CBL* tumour suppressor in myeloid neoplasms. *Nature* **460**, 904–908 (2009).
- Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40**, 722–729 (2008).
- Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467–472 (2011).
- Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465**, 473–477 (2010).
- Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
- Metzker, M. L. Sequencing technologies — the next generation. *Nature Rev. Genet.* **11**, 31–46 (2010).
- Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nature Biotechnol.* **26**, 1135–1145 (2008).
- Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
- Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469**, 539–542 (2011).
- Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).
- Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
- Yan, X. J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene *DNMT3A* in acute monocytic leukemia. *Nature Genet.* **43**, 309–315 (2011).

- Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–105 (2011).
- Nannya, Y. *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res.* **65**, 6071–6079 (2005).
- Yamamoto, G. *et al.* Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of Affymetrix single-nucleotide-polymorphism genotyping microarrays. *Am. J. Hum. Genet.* **81**, 114–126 (2007).
- Wahl, M. C., Will, C. L. & Luhrmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
- Tronchè, H., Wang, J. & Fu, X. D. A protein related to splicing factor *U2AF³⁵* that interacts with *U2AF⁵⁵* and SR proteins in splicing of pre-mRNA. *Nature* **388**, 397–400 (1997).
- Bevilacqua, L. *et al.* A population-specific *HTR2B* stop codon predisposes to severe impulsivity. *Nature* **468**, 1061–1066 (2010).
- Calvo, S. E. *et al.* High-throughput, pooled sequencing identifies mutations in *NUBPL* and *FOXRED1* in human complex I deficiency. *Nature Genet.* **42**, 851–858 (2010).
- Haase, D. *et al.* New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood* **110**, 4385–4395 (2007).
- Xiao, R. *et al.* Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. *Mol. Cell. Biol.* **27**, 5393–5402 (2007).
- Morin, R. D. *et al.* Somatic mutations altering *EZH2* (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nature Genet.* **42**, 181–185 (2010).
- Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Maquat, L. E. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nature Rev. Mol. Cell Biol.* **5**, 89–99 (2004).
- Ema, H. *et al.* Adult mouse hematopoietic stem cells: purification and single-cell assays. *Nature Protocols* **1**, 2979–2987 (2007).
- Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Rev. Mol. Cell Biol.* **10**, 741–754 (2009).
- Ni, J. Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).
- He, H. *et al.* Mutations in *U4atac* snRNA, a component of the minor spliceosome, in the developmental disorder MOPD I. *Science* **332**, 238–240 (2011).
- Edery, P. *et al.* Association of TALS developmental disorder with defect in minor splicing component *U4atac* snRNA. *Science* **332**, 240–243 (2011).
- David, C. J. & Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* **24**, 2343–2364 (2010).
- Pajares, M. J. *et al.* Alternative splicing: an emerging topic in molecular and clinical oncology. *Lancet Oncol.* **8**, 349–357 (2007).
- Shen, H., Zheng, X., Luecke, S. & Green, M. R. The *U2AF35*-related protein *Urp* contacts the 3' splice site to promote *U12*-type intron splicing and the second step of *U2*-type intron splicing. *Genes Dev.* **24**, 2389–2394 (2010).
- Barlow, J. L. *et al.* A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q- syndrome. *Nature Med.* **16**, 59–66 (2010).
- Ebert, B. L. *et al.* Identification of *RPS14* as a 5q- syndrome gene by RNA interference screen. *Nature* **451**, 335–339 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by Grant-in-Aids from the Ministry of Health, Labor and Welfare of Japan and from the Ministry of Education, Culture, Sports, Science and Technology, and also by the Japan Society for the Promotion of Science (JSPS) through the 'Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)', initiated by the Council for Science and Technology Policy (CSTP). pGCDNsmiRESEGF vector was a gift from M. Onodera. We thank Y. Mori, O. Hagiwara, M. Nakamura and N. Mizota for their technical assistance. We are also grateful to K. Ikeuchi and M. Ueda for their continuous encouragement throughout the study.

Author Contributions Y.Sh., Y.Sa., A.S.-O., Y.N., M.N., G.C., R.K. and S.Miyano were committed to bioinformatics analyses of resequencing data. M.Sa., A.S.-O. and Y.Sa. performed microarray experiments and their analyses. R.Y., T.Y., M.O., M.Sa., A.K., M.Sh. and H.N. were involved in the functional analyses of *U2AF35* mutants. N.O., M.S.-Y., K.I., H.M., W.-K.H., F.N., D.N., T.H., C.H., S.Miyawaki, S.C., H.P.K. and L.-Y.S. collected specimens and were also involved in planning the project. K.Y., Y.N., Y.Su., A.S.-O. and S.S. processed and analysed genetic materials, library preparation and sequencing. K.Y., M.Sa., Y.Sh., A.S.-O., Y. Sa. and S.O. generated figures and tables. S.O. led the entire project and wrote the manuscript. All authors participated in the discussion and interpretation of the data and the results.

Author Information Sequence data have been deposited in the DDBJ repository under accession number DRA000433. Microarray data have been deposited in the GEO database under accession numbers GSE31174 (for SNP arrays), GSE31171 (for exon arrays) and GSE31172 (for expression arrays). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.O. (sogawa-ty@umin.ac.jp).

Human oocytes reprogram somatic cells to a pluripotent state

Scott Noggle¹, Ho-Lim Fung², Athurva Gore², Hector Martinez¹, Kathleen Crumm Satriani^{3,4}, Robert Prosser^{3,4}, Kiboong Oum^{3,4}, Daniel Paull¹, Sarah Druckenmiller¹, Matthew Freeby^{5,6}, Ellen Greenberg^{5,6}, Kun Zhang², Robin Goland^{5,6}, Mark V. Sauer^{3,4}, Rudolph L. Leibel^{5,6} & Dieter Egli¹

The exchange of the oocyte's genome with the genome of a somatic cell, followed by the derivation of pluripotent stem cells, could enable the generation of specific cells affected in degenerative human diseases. Such cells, carrying the patient's genome, might be useful for cell replacement. Here we report that the development of human oocytes after genome exchange arrests at late cleavage stages in association with transcriptional abnormalities. In contrast, if the oocyte genome is not removed and the somatic cell genome is merely added, the resultant triploid cells develop to the blastocyst stage. Stem cell lines derived from these blastocysts differentiate into cell types of all three germ layers, and a pluripotent gene expression program is established on the genome derived from the somatic cell. This result demonstrates the feasibility of reprogramming human cells using oocytes and identifies removal of the oocyte genome as the primary cause of developmental failure after genome exchange.

The generation of animals by transfer of the genome from an adult cell into an unfertilized oocyte¹, and the isolation of pluripotent stem cells from human blastocysts², raised the prospect of generating stem cells with a patient's genome. This prospect holds much medical promise as these patient-specific stem cells could be used to generate differentiated cells for cell replacement. Unfortunately, progress towards this goal has been slowed by legal and social considerations limiting the availability of human oocytes for research. Despite these limitations, several studies were conducted^{3–11}, but none have achieved the derivation of a stem cell line. Thus, the question of whether human oocytes have the ability to reprogram somatic cells to a pluripotent state has remained unanswered.

Although it is now possible to induce pluripotent stem cell (iPS) formation by forced expression of transcription factors in somatic cells¹², differences between iPS- and blastocyst-derived stem cells have been reported for gene expression^{13,14}, DNA methylation^{15,16} and differentiation potential¹⁷. In addition, reprogramming to iPS cells seems to compromise genomic integrity, introducing *de novo* mutations¹⁸ and copy number variations^{19,20}. Whether reprogramming using human oocytes yields pluripotent stem cells without these abnormalities remains to be determined.

Various sources have been explored, including failed fertilized oocytes⁴, oocytes deemed in excess of clinical need^{3,5,21}, *in vitro* matured oocytes¹⁰ and fertilized oocytes²². Previously, we have found that very few women agree to donate their oocytes for research without payment¹¹. The majority of oocyte donors believe that payment should be provided regardless of whether the oocytes are used for research or reproductive purposes²³. Payment for reproductive oocyte donation is common in the USA with more than 8,000 donor *in vitro* fertilization (IVF) cycles performed annually²⁴. Recognizing the varying views on payments to research oocyte donors, the American Society for Reproductive Medicine and the International Society for Stem Cell Research have proposed balanced guidelines^{25,26} which allow payment at the discretion of research oversight committees, which must ensure

that financial considerations do not constitute an undue inducement. Following on from those guidelines, we have developed protocols that were reviewed and approved by the institutional review board and stem cell committees of Columbia University. These protocols allowed women participating in the reproductive egg donation program to select between donation for reproductive purposes and donation for research, offering equal remuneration regardless of their choice. Consequently, the decision to donate was before and independent of their decision to donate for research. Our study of 270 mature human oocytes revealed that the exchange of the oocyte genome with the genome of a somatic cell consistently leads to developmental arrest. However, when the oocyte genome is not removed, and the somatic cell genome is merely added, the activated human oocytes develop to the blastocyst stage. Human stem cells derived from these blastocysts contain both a haploid genome derived from the oocyte and a diploid somatic cell genome reprogrammed to a pluripotent state.

Results

Development fails after somatic genome replacement

In several mammalian species, somatic cell reprogramming has been achieved by replacing the oocyte genome at metaphase II (MII) of meiosis with a somatic cell nucleus (Fig. 1a). To remove the oocyte genome, we identified the location of the spindle-chromosome complex in 38/50 MII oocytes (Fig. 1b). All oocytes (43/43) survived genome removal with (31) or without (12) addition of Hoechst stain/minimal ultraviolet light exposure to verify enucleation. Oocytes lacking a genome were used for the transfer of somatic cell genomes obtained from skin cells of a male diabetic (T1D) and a healthy male adult. They were labelled with a green fluorescent protein (GFP) or a histone 2b (H2B):GFP transgene under the control of the ubiquitously expressed CAGGS promoter (Fig. 1c). Because Hoechst staining seemed to inhibit nuclear remodelling in rhesus oocytes²⁷, we monitored chromosome condensation every hour after transfer. All oocytes (35/35), whether or not they had been exposed to Hoechst, condensed the

¹The New York Stem Cell Foundation Laboratory, New York, New York, USA. ²Department of Bioengineering, University of California at San Diego, La Jolla, California, USA. ³Center for Women's Reproductive Care, College of Physicians and Surgeons, Columbia University, New York, New York, USA. ⁴Department of Obstetrics and Gynecology, College of Physicians and Surgeons, Columbia University, New York, New York, USA. ⁵Naomi Berrie Diabetes Center, College of Physicians and Surgeons, Columbia University, New York, New York, USA. ⁶Department of Pediatrics, College of Physicians and Surgeons, Columbia University, New York, New York, USA.

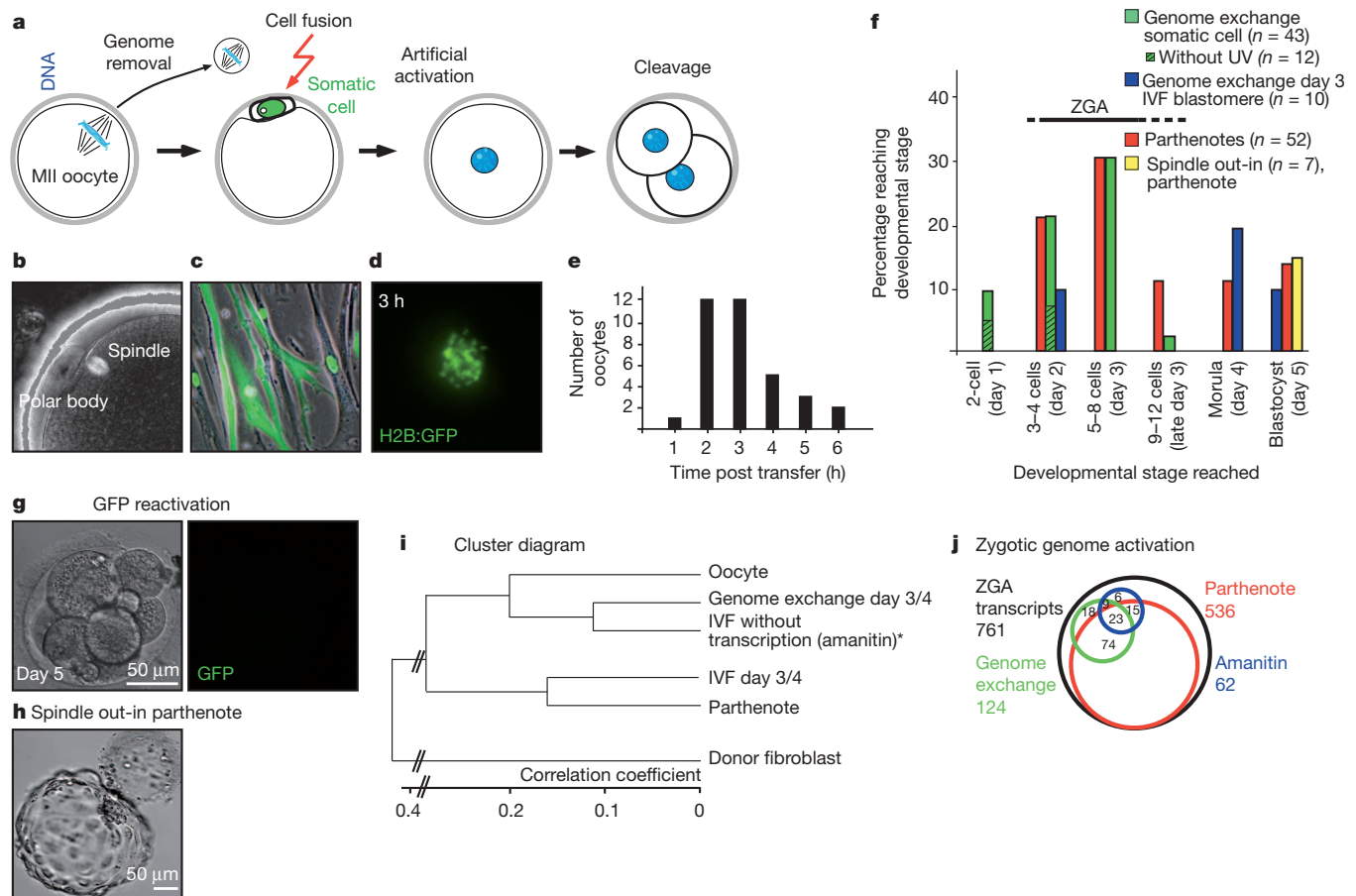


Figure 1 | Developmental and transcriptional defects after genome exchange. **a**, Schematic of genome exchange in human oocytes. **b**, Human oocyte at the MII stage, viewed by microtubule birefringence. **c**, Donor cell population marked with either H2B:GFP or GFP. **d**, Somatic chromatin 3 h after transfer. **e**, Timing of chromosome condensation. **f**, Developmental potential. Vertical axis is the percentage of activated eggs reaching specific developmental stages (horizontal axis). Days indicate the time points of normal

somatic chromosomes (Fig. 1d, e); upon activation, 22/31 (71%) of the oocytes continued normal cleavage development. However, as we had previously observed following nuclear transfer into human zygotes²², development arrested at a stage of 6–10 cells (Fig. 1f, g).

As a control for the quality of the oocytes, the development of *in vitro* fertilized donor oocytes was followed at the IVF clinic; 16/21 (76%) developed to the blastocyst or morula stage by day 6 of culture, indicating excellent developmental potential (Supplementary Table 1). Likewise, artificially activated oocytes developed to the morula and blastocyst stages (13/52 or 25%), well beyond the point of developmental arrest seen after genome exchange (Fig. 1f).

As a control for our experimental manipulations, we transferred H2B:GFP-labelled somatic nuclei without immediately removing the oocyte genome (Supplementary Fig. 1a). Six to eight hours after artificial activation two interphase nuclei had formed within a single cell (Supplementary Fig. 1b). The H2B:GFP-labelled genome could be distinguished from the unlabelled oocyte genome, and either of them specifically extracted using contrast optics and GFP fluorescence (Supplementary Fig. 1c, e); Hoechst staining and ultraviolet illumination was not required. Both types of cells therefore experienced the same manipulations, but differed in their ultimate genetic content, having either the oocyte genome or the somatic cell genome. Activated oocytes containing only the somatic genome formed 4–12 cells, but all arrested without reactivating the GFP transgene (32/32) (Supplementary Fig. 1d). In contrast, activated oocytes containing only the oocyte genome cleaved and 4/7 (57%) developed to the

developmental progression. UV, ultraviolet light. **g**, Arrested development after somatic genome exchange. **h**, Development after spindle removal and re-transfer. **i**, Cluster diagram of global gene expression. *from ref. 22. **j**, Venn diagram of transcripts elevated in IVF samples on day 3–4 of development (black circle) in comparison to oocytes. The overlap with parthenotes, amanitin-treated samples and genome exchange samples are shown.

blastocyst stage (Supplementary Fig. 1f), allowing the generation of pluripotent parthenogenetic stem cells (Supplementary Fig. 2).

To test whether developmental potential was determined by the state of differentiation of the transferred genome, we replaced the oocyte genome with that of a blastomere. Upon activation, development to the morula and blastocyst stage occurred (Fig. 1f and Supplementary Fig. 3). Furthermore, following the removal of the oocyte's genome and subsequent retransfer into the same oocyte, parthenogenetic development to the blastocyst stage was observed (Fig. 1f, h and Supplementary Fig. 4).

Development arrests with transcriptional defects

Extensive transcriptional activity from the zygotic genome normally starts at the 4–8 cell stage²⁸, or on day 3 of development, coincident with the stage of developmental arrest after somatic genome transfer (Fig. 1f). To determine whether the somatic genome was being expressed, we compared the transcriptome at the 6–12 cell stage after genome exchange, to the transcriptome after artificial activation on day 3/early day 4 of development. To distinguish between expression from the transferred genome and maternal contributions to transcript abundance, we compared our data to those obtained after development of fertilized eggs from the 1-cell to the 6–8-cell stage in media containing the RNA polymerase II inhibitor alpha amanitin²². Using hierarchical clustering of the global gene expression patterns, we found that after genome exchange, transcript types and abundances at the 6–12-cell stage most closely resembled a state of inhibition of transcriptional activity (Fig. 1i).

We then identified transcripts that were relatively upregulated in comparison to unfertilized oocytes. Using day 3 and day 4 IVF controls, we defined 761 transcripts upregulated at zygotic genome activation (more than fivefold, $P < 0.01$). Of these 761 transcripts, only 124 (16%) were upregulated after genome exchange, and 62 (8%) were upregulated after amanitin treatment (more than fivefold, $P < 0.01$), possibly reflecting differential mRNA stability (Fig. 1j). In contrast, in parthenotes 536/761 transcripts (70%) ($P < 0.001$) were more than fivefold upregulated.

To determine if the developmental arrest correlated with continued expression of somatic cell genes, we identified 1,406 genes that were expressed at higher levels in the skin donor fibroblasts than in MII oocytes (more than tenfold, $P < 0.001$ for four biological replicates). The average transcript levels of these genes after genome transfer was 0.4-fold lower than in IVF controls. Therefore, neither the transcriptional program of a somatic cell nor that of a blastomere was being expressed. Among the few genes that were specifically elevated after genome exchange (Supplementary Table 2), one, *GADD45G*, is involved in stress-induced cell cycle arrest.

The oocyte genome rescues development

These developmental defects could be caused by an inability of the somatic cell genome to be appropriately expressed, replicated and/or segregated during cleavage development. Alternatively, molecules specific to the oocyte genome for which the somatic nucleus is unable to compensate may be removed during oocyte genome removal.

To distinguish between these possibilities, we transferred a somatic cell genome but did not remove the oocyte genome (Fig. 2a). In contrast to previous experiments, development continued to the compacted morula stage, and expression of the *CAGGS:GFP* transgene was re-initiated at the appropriate stage (35/35 cleavage stages with four or more cells). Development to the blastocyst stage was efficient (13 blastocysts of 63 transferred oocytes, or 21%), indicating that the somatic cell genome did not interfere with development to the blastocyst stage.

Derivation of pluripotent stem cells

From these blastocysts, we isolated the inner cell mass and derived two cell lines, soPS1 (for somatic cell genome, oocyte genome pluripotent stem cell 1), containing the genome of a male T1D subject, and soPS2, containing the genome of a healthy male adult. Both cell lines were triploid (Fig. 3a), containing short tandem repeat (STR) alleles consistent with the presence of a diploid somatic cell genome and the haploid genome of the oocyte (Supplementary Tables 3 and 4). soPS1 contained an additional chromosome 17 of somatic donor cell origin and a balanced translocation between chromosomes 15 and 17 (Supplementary Fig. 5a, b). At passage 23, 30% of soPS1 cells had gained additional copies of chromosome 12 and 17, chromosomal aberrations that commonly occur in pluripotent stem cell cultures because they confer a growth advantage²⁹. soPS2 was karyotypically stable over more than 20 passages (Supplementary Material). During a period of 6 months, soPS1 and soPS2 completed more than 30 passages or over 100 population doublings (Supplementary Fig. 5c) without undergoing replicative crisis. Mitochondrial genomes were of oocyte donor origin without sign of heteroplasmy in either cell line (Fig. 3b and Supplementary Fig. 5d). Mitochondria transferred with the somatic nucleus may be outnumbered by the mitochondria of the oocyte, or they may be lost during cleavage development.

Both soPS cell lines expressed molecular markers characteristic of pluripotent stem cells (Fig. 3c), and when differentiated *in vitro*, or following injection into immunocompromised mice, cell types representative of all three germ layers were observed (Fig. 3d). The global gene expression profile of both soPS cell lines clustered closely with that of other pluripotent cell types, including NYSCF1, a stem cell line derived from an IVF blastocyst. The parthenogenetic stem cell line, pPS1, and iPS cell lines derived from both skin cell donors also clustered closely with soPS cells, but the somatic donor cells clustered separately (Fig. 3e). We identified 1,327 genes that were differentially expressed between soPS2 and its donor fibroblast ($P < 0.01$). Of these

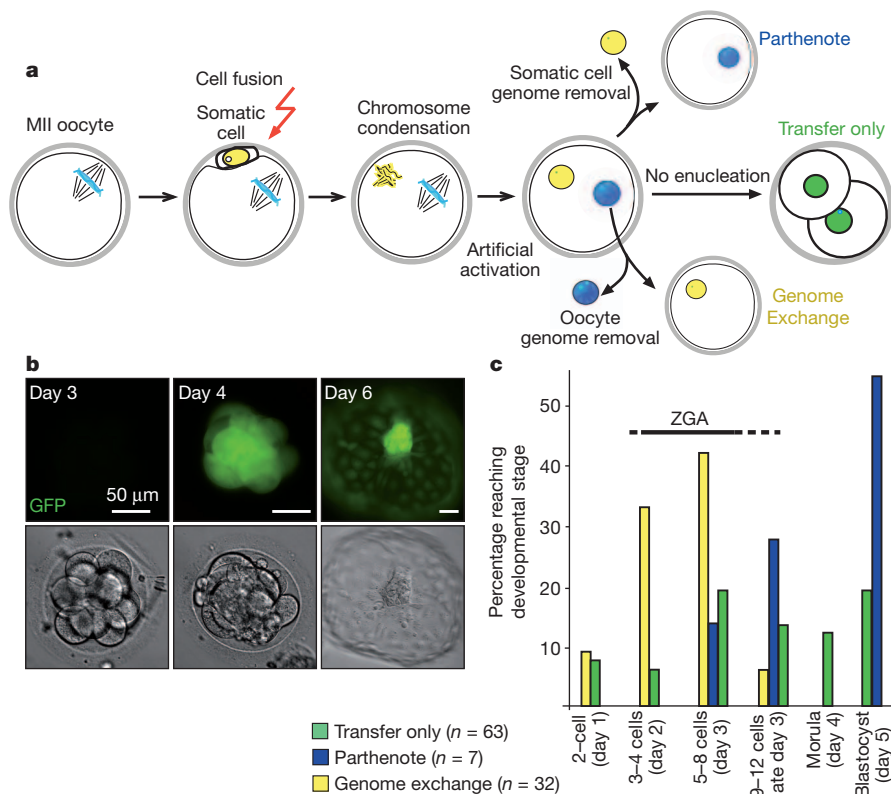


Figure 2 | Development after somatic cell genome transfer with retention of the oocyte genome. **a**, Schematic of somatic genome transfer without or with removal of either the oocyte or the somatic cell genome at the first interphase.

b, Developmental progression 'transfer only'. Days post artificial activation are indicated. **c**, Developmental potential. Vertical axis is the percentage of activated eggs reaching specific developmental stages (horizontal axis).

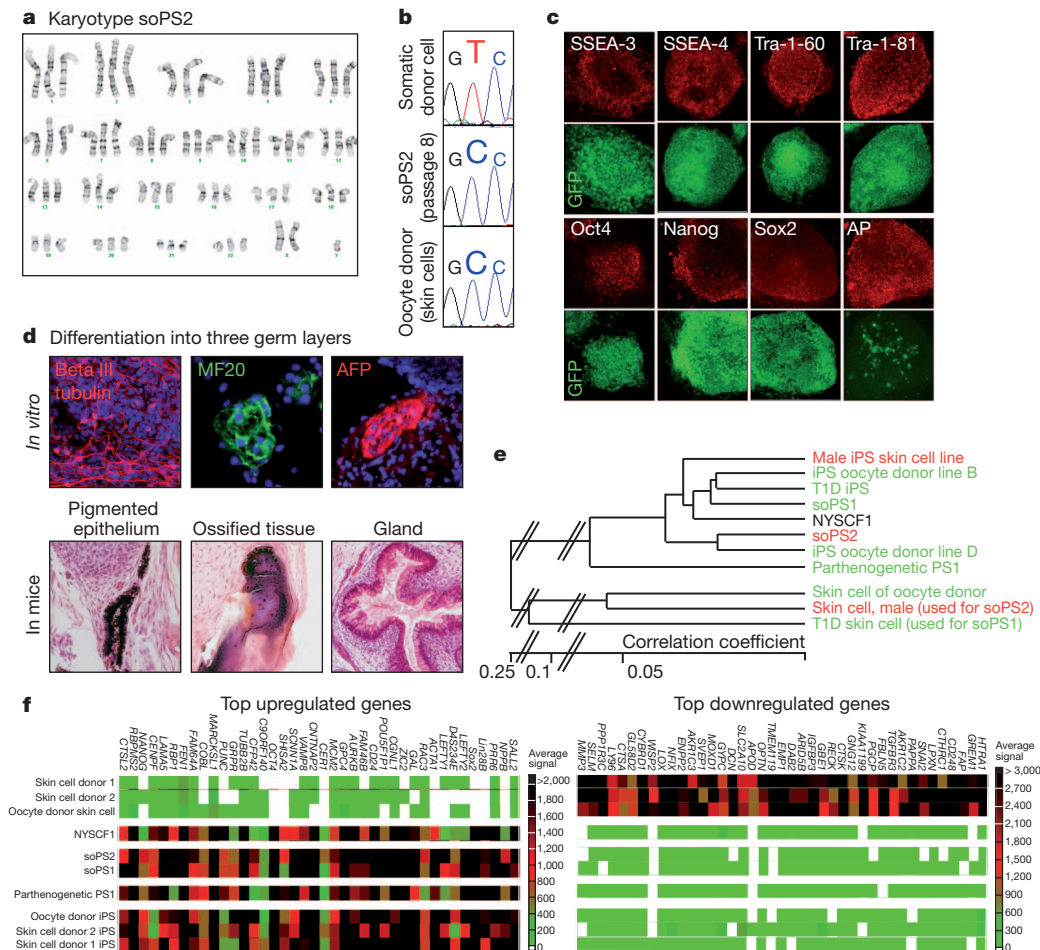


Figure 3 | soPS cells are pluripotent. **a**, Karyotype. **b**, Sequence of the mitochondrial hypervariable region I. **c**, Immunostaining for pluripotency markers. **d**, Immunostaining and histochemistry of cells/tissues upon differentiation. **e**, Cluster diagram of global gene expression analysis. Cell lines

1,327 transcripts, 463 were present at fivefold higher levels in the stem cells than in the fibroblasts, and 670 transcripts were decreased by a factor of five or more in soPS2. Among the genes with the most significant upregulation in soPS1 and soPS2 were genes typically expressed in pluripotent stem cells, but not in fibroblasts, such as *LIN28A*, *POU5F1*, *SOX2*, *NANOG* and *LEFTY1* (Fig. 3f). Genes that were most downregulated included those typically expressed in fibroblasts, such as fibroblast activating protein (*FAP*), pappalysin (*PAPP*), metalloproteinase (*MMP3*), a collagen triple helix-containing protein (*CTHRC1*), and a mesoderm-specific transcription factor (*SNAI2*) (Fig. 3f). In comparison to NYSCF1, 28 genes and 24 genes were expressed at higher levels in soPS2, and soPS1 respectively (more than threefold, $P < 0.01$) (Supplementary Table 5), including neuropeptide galanin, neuronatin, *SRY*, *rex1* (also known as *ZFP42*), *NODAL*, *Cerberus 1* and *LEFTY2*. Presumably, expression of these genes reflects spontaneous differentiation into various cellular lineages in soPS cultures, rather than incomplete reprogramming of the somatic cell genome. In additional comparisons we were not able to identify consistent differences between soPS cells, iPS cells, NYSCF1 and pPS1 (Supplementary Fig. 6).

Epigenetic reprogramming

Consistent with reprogramming of the somatic cell genome to a pluripotent state, methylation of DNA at the *Nanog* promoter was low (5–15%) in soPS cells and high (38–58%) in the somatic donor cells (Fig. 4a). Demethylation at the *Nanog* and *Oct4* promoters correlated with the expression of single nucleotide polymorphisms (SNPs) located in the somatic genome (Fig. 4b). Demethylation was specific

of the same colour are genetically related to either soPS1 or soPS2. **f**, Heat map of most highly up- or downregulated genes in stem cell lines in comparison to the donor fibroblasts.

and did not occur on an imprinted locus *PEG3*: two thirds of sequencing reads were methylated in soPS1, reflecting the presence of two methylated maternal alleles, one from the oocyte and one from the somatic cell, as well as a single paternal allele of somatic origin (Fig. 4c).

To determine whether reprogramming had occurred at other loci, we used a genome-wide digital allelotyping approach to distinguish gene expression from the somatic cell-derived genome and the oocyte-derived genome in soPS cells. This method is based on a library of 27,000 'padlock' probes flanking known SNPs on all 23 chromosomes of the human genome (Fig. 4d)³⁰. Extension of the padlock probes by DNA polymerase captures the SNP and allows single molecule DNA sequencing. SNP capture on genomic DNA will reflect the allelic ratio of the SNP, whereas SNP capture on cDNA will reflect the transcriptional activity of an allele.

We prepared genomic DNA and cDNA from soPS cells and their corresponding somatic cells (Supplementary Fig. 7). By generating 108,982,981 sequencing reads (Supplementary Table 6), we were able to identify 787 and 483 expressed SNPs for soPS1 and soPS2, respectively, for which the oocyte donor DNA sequence differed from that of the somatic cell and for which the somatic cell was homozygous. The median allelic ratio (somatic/(oocyte plus somatic)) for the genomic DNA was 0.64 for both soPS1 and soPS2. This ratio is consistent with the inference that a diploid complement of 46,XY chromosomes originating from the somatic cell and a haploid set of 23,X chromosomes originating from the oocyte are present in the soPS cell lines.

To calculate the proportion of transcripts expressed from the somatic cell-derived genome in soPS cells (cDNA somatic/(cDNA oocyte plus cDNA somatic)), each individual SNP was normalized to the ratio

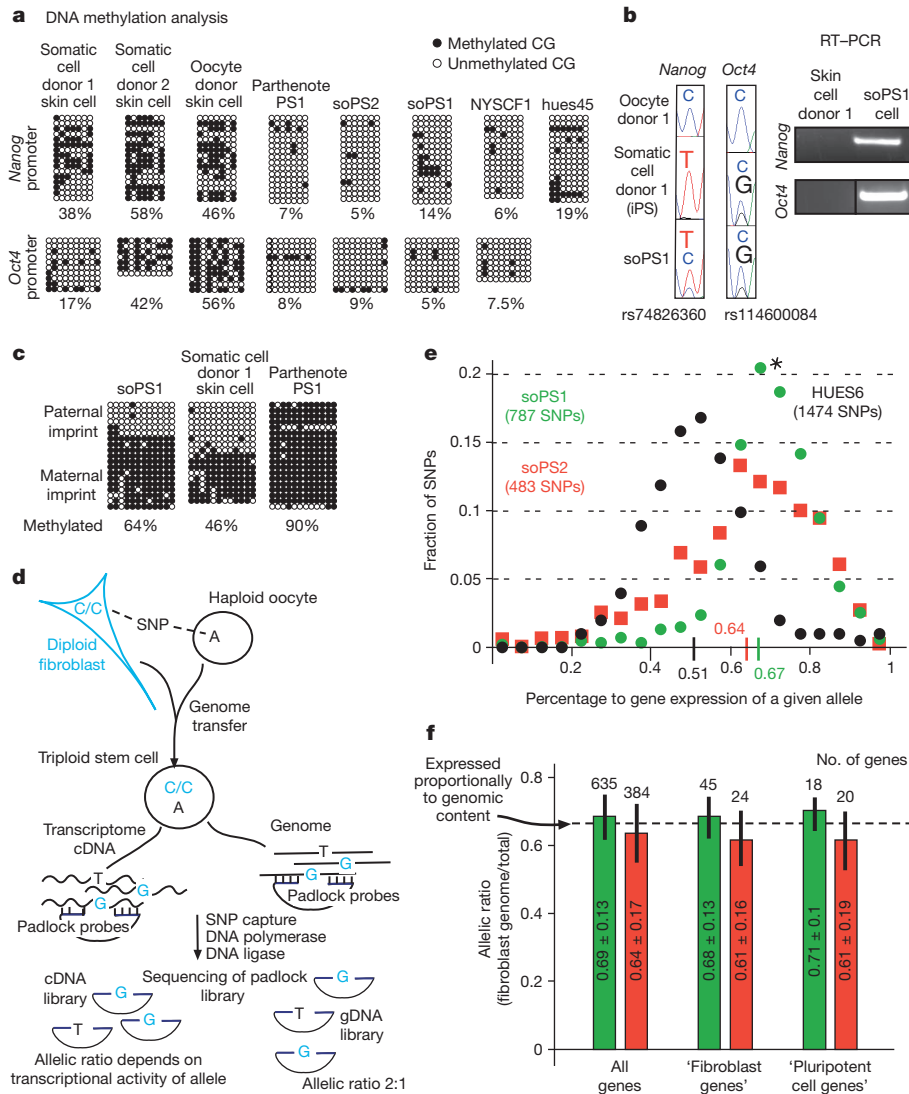


Figure 4 | Human oocytes reprogram a somatic cell to a pluripotent state. **a**, Bisulphite sequencing of the *Nanog* and *Oct4* promoters. The percentage of methylated CG is shown. **b**, PCR with reverse transcription (RT-PCR) and sequencing of expressed SNPs. **c**, Bisulphite sequencing of the differentially methylated region (DMR) of *PEG3*. **d**, Schematic of genome-wide SNP capture and allelotyping. **e**, The reprogrammed genome contributes two thirds of the transcriptome. Alleles expressed at a certain allelic ratio (binned in intervals of 0.05, horizontal axis) displayed as fraction of all captured alleles (vertical axis). **f**, No memory of somatic gene expression is observed. Mean allelic ratios and standard deviations in soPS cells for all captured SNPs represented in the Illumina array 'all genes', for SNPs in genes highly expressed in the fibroblast donor cell ('fibroblast genes'), and for SNPs in genes highly expressed in pluripotent stem cells ('pluripotent cell genes'). soPS1, green bars; soPS2, red bars. Mean \pm s.d. is shown.

of the same SNP observed in genomic DNA. If a locus was expressed in proportion to its genomic content in soPS cells, the ratio of transcripts would be expected to equal $2/3 = 0.6667$. Allele ratios for each SNP were binned in increments of 0.05 units, and that number of transcripts was expressed as a fraction of total SNPs captured (Fig. 4e). (For example, 162 somatic cell SNPs were expressed at a ratio of 0.65–0.7 in soPS1. As the total number of SNPs captured was 787, the fraction is $162/787 = 0.206$, yielding the data point indicated by an asterisk in Fig. 4e). The median of the allelic ratio was 0.67 for soPS1 and 0.64 for soPS2, consistent with expression from the somatic genome proportional to the genomic content. The distribution of allelic ratios approximated a Gaussian curve (Shapiro–Wilk test for normality $W = 0.92$ for soPS1 and $W = 0.97$ for soPS2), indicating significant variability in the contribution of a particular allele. Such variation was not specific to soPS cells. Variability at comparable levels was also observed for allelic ratios in diploid HUES6 as well as in tetraploid HUES6–somatic cell hybrids³¹ (Supplementary Fig. 8), and may be caused by polymorphisms in gene regulatory regions³⁰.

Our expectation was that if reprogramming in soPS cells were incomplete we would detect a bias in this distribution: genes expressed at high levels in fibroblasts, but at low levels in pluripotent stem cells, would be expressed predominantly from the fibroblast cell-derived genome in soPS cells; conversely, genes that are expressed at high levels in pluripotent stem cells, but not in fibroblasts, would be expressed predominantly from the oocyte-derived genome. Among

a total of 1,019 genes represented in both gene expression array as well as in the SNP capture data, 38 (18 for soPS1 and 20 for soPS2) were upregulated at least fourfold in soPS cells ($P < 0.01$), and 69 (45 for soPS1 and 24 for soPS2) genes were expressed in the donor fibroblasts at levels fivefold or higher compared to soPS cells ($P < 0.01$). The mean allelic transcript ratio for the 'somatic cell genes' and the 'pluripotent cell genes', was identical, and did not differ from the expected allelic ratio of 0.66 nor from the allelic ratio of all captured genes (Fig. 4f and Supplementary Table 7).

Discussion

Upon replacement of the oocyte genome with that of a somatic cell, we observed developmental arrest at late cleavage stages in association with severe transcriptional abnormalities, similar to the arrest we had previously observed following somatic cell genome exchange in human zygotes²². These defects occurred despite the use of high quality oocytes obtained from women without history of infertility. This result is consistent with previous studies^{4,9,10}, but contrasts with a report of efficient development to the blastocyst stage³. Those authors attributed the improved developmental potential to oocyte quality. However, because the STR genotype of the blastocysts was incomplete, an alternative interpretation is that the presence of genetic material of the oocyte promoted development to the blastocyst stage. Another group has generated a single blastocyst after transfer of pluripotent stem cell genomes⁵, suggesting that the developmental arrest seen with

somatic cells may not apply to pluripotent cells. Consistent with this hypothesis, we find that after transfer of blastomere or oocyte nuclei, development to the blastocyst stage occurs.

In contrast, if the somatic cell genome is merely added and the oocyte genome is not removed, development to the blastocyst stage occurs. From these blastocysts, we were able to derive triploid pluripotent stem cells containing a diploid genome complement of the somatic cell and a haploid genome complement of the oocyte.

It was previously shown that epigenetic memory is often retained in mouse iPS cells, but not in cells reprogrammed by mouse oocytes³². We compared gene expression from the haploid oocyte genome, which reached pluripotency through its developmental trajectory, with gene expression from the diploid somatic genome, which required a reprogramming process to establish pluripotency. Because both of these genomes are present in the same cell, they are exposed to an identical environment. Preferential expression of pluripotency genes from the oocyte genome, or preferential expression of somatic genes from the somatic cell genome, would be a strong indication of epigenetic memory. Using a genome-wide allelotyping approach and gene expression profiling, we compared the levels of expression from each allele, and did not find evidence for epigenetic memory: expression from the reprogrammed somatic genome was proportional to the genomic content and did not depend on the activity in the fibroblast donor cell.

This report demonstrates the feasibility of somatic cell reprogramming using human oocytes. With a reliable source of human oocytes, it should be possible to overcome the requirement of the oocyte genome for somatic cell reprogramming, allowing the generation of diploid pluripotent stem cells.

METHODS SUMMARY

Human oocytes were aspirated approximately 36 h after human chorionic gonadotropin application and transported to the laboratory in a portable incubator at 37 °C. The oocyte genome of the MII oocyte was identified by microtubule birefringence and/or staining in Hoechst 33342 and minimal ultraviolet illumination, whereas in activated oocytes, the oocyte genome was identified by Hoffmann modulation contrast optics only and then removed by laser-assisted micromanipulation. Somatic cells were introduced into the oocyte. Oocytes were activated in the calcium ionophore ionomycin, followed by incubation in the kinase inhibitor 6-dimethylaminopurine, thoroughly washed, and allowed to develop.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 April; accepted 1 August 2011.

- Wilmot, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997).
- Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
- French, A. J. *et al.* Development of human cloned blastocysts following somatic cell nuclear transfer with adult fibroblasts. *Stem Cells* **26**, 485–493 (2008).
- Hall, V. J. *et al.* Developmental competence of human *in vitro* aged oocytes as host cells for nuclear transfer. *Hum. Reprod.* **22**, 52–62 (2007).
- Stojkovic, M. *et al.* Derivation of a human blastocyst after heterologous nuclear transfer to donated oocytes. *Reprod. Biomed. Online* **11**, 226–231 (2005).
- Cibelli, J. *et al.* Somatic cell nuclear transfer in humans: pronuclear and early embryonic development. *J. Regen. Med.* **2**, 25–31 (2001).
- Kennedy, D. Editorial retraction. *Science* **311**, 335 (2006).
- McElroy, S. L. *et al.* Developmental competence of immature and failed/abnormally fertilized human oocytes in nuclear transfer. *Reprod. Biomed. Online* **16**, 684–693 (2008).
- Chung, Y. *et al.* Reprogramming of human somatic cells using human and animal oocytes. *Cloning Stem Cells* **11**, 213–223 (2009).
- Heindryckx, B., De Sutter, P., Gerris, J., Dhont, M. & Van der Elst, J. Embryo development after successful somatic cell nuclear transfer to *in vitro* matured human germinal vesicle oocytes. *Hum. Reprod.* **22**, 1982–1990 (2007).
- Egli, D., Chen, A. E., Melton, D. & Eggan, K. Impracticability of egg donor recruitment in the absence of compensation. *Cell Stem Cell* doi:10.1016/j.stem.2011.08.002 (in the press).

- Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Chin, M. H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111–123 (2009).
- Ghosh, Z. *et al.* Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. *PLoS ONE* **5**, e8975 (2010).
- Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genet.* **41**, 1350–1353 (2009).
- Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68–73 (2011).
- Hu, B. Y. *et al.* Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proc. Natl Acad. Sci. USA* **107**, 4335–4340 (2010).
- Gore, A. *et al.* Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63–67 (2011).
- Hussein, S. M. *et al.* Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58–62 (2011).
- Mayshar, Y. *et al.* Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* **7**, 521–531 (2010).
- Revazova, E. S. *et al.* Patient-specific stem cell lines derived from human parthenogenetic blastocysts. *Cloning Stem Cells* **9**, 432–449 (2007).
- Egli, D., Chen, A. E., Melton, D. & Eggan, K. Reprogramming occurs within hours after mouse but not human nuclear transfer. *Nature Commun.* doi:10.1038/ncomms1503 (in the press).
- Klitzman, R. & Sauer, M. V. Payment of egg donors in stem cell research in the USA. *Reprod. Biomed. Online* **18**, 603–608 (2009).
- Society for Assisted Reproductive Technology & the American Society for Reproductive Medicine. Assisted reproductive technology in the United States: 2001 results generated from the American Society for Reproductive Medicine/Society for Assisted Reproductive Technology registry. *Fertil. Steril.* **87**, 1253–1266 (2007).
- The Ethics Committee of the American Society for Reproductive Medicine. Financial compensation of oocyte donors. *Fertil. Steril.* **88**, 305–309 (2007).
- Daley, G. Q. *et al.* Ethics. The ISSCR guidelines for human embryonic stem cell research. *Science* **315**, 603–604 (2007).
- Mitalipov, S. M. *et al.* Reprogramming following somatic cell nuclear transfer in primates is dependent upon nuclear remodeling. *Hum. Reprod.* **22**, 2232–2242 (2007).
- Braude, P., Bolton, V. & Moore, S. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* **332**, 459–461 (1988).
- Draper, J. S. *et al.* Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nature Biotechnol.* **22**, 53–54 (2004).
- Zhang, K. *et al.* Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nature Methods* **6**, 613–618 (2009).
- Cowan, C. A., Atienza, J., Melton, D. A. & Eggan, K. Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. *Science* **309**, 1369–1373 (2005).
- Kim, K. *et al.* Epigenetic memory in induced pluripotent stem cells. *Nature* **467**, 285–290 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank our research subjects for participating. We thank S. Solomon and K. Eggan for discussions and support, L. Bauer for help with blastocyst thawing, D. Kahler for cell sorting, M. Verbitsky and S. Kisselev for microarray hybridization, V. Miljkovic for Affymetrix SNP chip hybridization, C. LeDuc and Y. Ravussin for help with data analysis and mouse work, R. Maehr for DiPS H.1.5, C. Marshall and J. Safran for administrative support, S. Paull for cover art, Z. Hall and S. Chang for critical reading of the manuscript. This research was supported by a UCSD startup fund to K.Z., the New York Stem Cell Foundation primarily, and the Russell Berrie Foundation.

Author Contributions R.G. and M.V.S. wrote IRB and consent documents, M.V.S., K.C.S., K.O. and R.P. consented oocyte donors and retrieved oocytes, D.E. and S.N. designed and performed experiments with oocytes, D.E., H.-L.F., A.G., H.M., D.P. and K.Z. characterized stem cell lines, M.F., E.G. and M.V.S. performed skin biopsies, D.E. performed skin cell isolation, soPS and iPS derivation, S.D. performed NYSCF1 derivation, D.E. and R.L.L. wrote the paper with input from all authors. All work with human oocytes and stem cells was performed at the NYSCF laboratory.

Author Information Illumina array data have been deposited at GEO under accession number GSE28024. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.E. (d.egli@nyscf.org).

METHODS

Oocyte donation. Oocyte donors of age 22–33 were recruited from the women participating in the reproductive oocyte donation program at the Center for Women's Reproductive Care (CWRC) at Columbia University P&S. These women had made a decision to enter the reproductive egg donation program, they met all criteria required for donation for reproductive purposes, and only then were presented with the option to donate oocytes for research. Both licensed medical social workers and CWRC physicians screened all women with respect to their reproductive, medical and psychosocial health. All of the women had a college degree or additional higher education, and none were financially disadvantaged. All women in the study were fully employed. During a period of 19 months, 16 women out of the 252 women enrolled in the reproductive oocyte donation program were asked if they wanted to donate oocytes to research. These women discussed the stem cell study in detail with a physician and those who chose to donate oocytes to research gave signed informed consent and initiated a standard hormone control regimen. All 16 women decided to participate in the study and gave informed consent (100% compliance). Two women did not complete the hormone treatment because of a lack of response. Two additional women donated for the study at a later time. In total, 16 women donated 270 mature MII oocytes (range of 2–26, or a mean of 16.9 oocytes per donor cycle). Payment for participation was equal to payment for women donating oocytes for reproduction at CWRC, or \$8,000 (pre-tax).

Skin biopsies. Skin biopsies (3 mm) were obtained using an AcuPunch biopsy Kit (Acuderm Inc.) from the locally anesthetized (1% Lidocaine HCl, Hospira, Inc.) upper arm or the upper leg. Biopsies were cut in 10–15 smaller pieces, placed in a six-well dish around a droplet of silicon grease, covered with a glass cover slip, and allowed to grow for 3–4 weeks in medium containing DMEM, 10% FBS, 1% Anti-Anti, nucleosides, GlutaMAX, β -mercaptoethanol and nonessential amino acids (all Invitrogen). In some instances, skin biopsies were obtained from subjects also donating oocytes. The identification numbers of the skin cell donors are 1-000 (male, T1D used for generation of soPS1) and 1-016 (male, used for generation of soPS2), and 1-034 for an oocyte donor. Protocols for obtaining skin biopsies and for their use in reprogramming experiments were reviewed and approved by the institutional review board and stem cell committees of Columbia University. All subjects gave signed informed consent.

Genome transfer into human oocytes and stem cell derivation. Oocytes were transported in GMOPsplus (Vitrolife) in a portable incubator (INC-RB1, CryoLogic) at 37 °C. The oocyte genome was identified by microtubule birefringence using the Oosight imaging system, and/or staining in 2 μ g ml⁻¹ Hoechst 33342 and minimal ultraviolet illumination. All manipulations were done on a Nikon TE2000-U equipped with Narishige micromanipulators and a Tokai hit heating plate. Somatic cells were infected with a vesicular stomatitis virus G protein (VSVG)-pseudotyped CAGGS:GFP or CAGGS:H2B-GFP retrovirus, sorted for GFP expression with a BD FACSAriaIIu, and grown to confluence to induce cell cycle exit. A single somatic cell was inserted below the zona pellucida of the oocyte using laser-assisted zona drilling (Hamilton Thorne) and introduced into the oocyte either by two fusion pulses of 20 μ s width and 1.3 kV cm⁻¹ strength (LF201, NEPA Gene), in cell fusion medium 0.26 M mannitol, 0.1 mM MgSO₄, 0.05% BSA, 0.5 mM HEPES, or by prior incubation of the somatic cell in inactivated Sendai virus HVJ-E (GenomeOne, Cosmo Bio), diluted with fusion buffer 1:5. For both fusion methods, the efficiency of fusion and oocyte survival was close to 100%. The first polar body was removed or ablated with two to three 500- μ s laser pulses to avoid potential fusion to the oocyte. Oocytes were activated in 5 μ M ionomycin (Sigma) in GMOPs plus for 5 min, followed by 4–5 h incubation in 2 mM 6-DMAP (Sigma) or until small interphase nuclei became apparent, thoroughly washed in Global medium and cultured to the blastocyst stage at 37 °C (Minc incubator, Cook), in a certified gas mixture 5% O₂, 6% CO₂, 89% N₂ (TechAir). Some samples were harvested on day 3–5 of development for gene expression analysis. Blastocysts were used for derivation of pluripotent stem cells as described³³, with the addition of 2 μ M Thiazovivin (Stemgent) and 10 μ M Rock-Inhibitor Y-27632 (Stemgent) to the derivation medium. Human pluripotent stem cells were expanded manually or enzymatically and cultured under standard conditions, as previously described³⁴. Human blastocysts for the derivation of NYSCF1 and human cleavage stages for gene expression analysis were thawed using the Sidney IVF thawing kit (K-SITS-5000, Cook Medical). Human blastocysts and cleavage stages were obtained from anonymous donors at CWRC under protocols reviewed and approved by the Columbia stem cell committee and the Columbia IRB. NYSCF1 characterization is described in Supplementary Fig. 9. iPS cells were generated according to published protocols using VSVG-pseudotyped retroviruses³⁵.

Gene expression analysis. RNA from human cleavage stages and blastocysts was isolated using a picopure RNA isolation kit (Arcturus), amplified by two rounds of T7 transcription using the total Prep RNA amplification kit (Illumina). RNA

from cell lines was isolated using RNeasy Plus mini kit (Qiagen) and amplified with a single round of T7 transcription. Amplified biotin labelled RNA was hybridized to the Illumina HumanRef-8 v3 Expression BeadChips. Analysis was undertaken using GenomeStudio and Microsoft Excel programs as follows: data were normalized to the average signal. Background was subtracted. Unfertilized oocytes (two biological replicates consisting of five MII oocytes) were used as a reference point for all comparisons. Transcripts (1,345) were more than fivefold upregulated ($P < 0.01$) in IVF controls on day 3 (two biological replicates consisting of two specimens). Of these transcripts 761 were also elevated in IVF controls collected early on day 4 (two biological replicates consisting of 17 specimen, more than fivefold, $P < 0.01$). These were defined as the ZGA transcripts. We then determined how many of these ZGA transcripts were also upregulated after genome exchange (three biological replicates, nine specimens of up to 12 blastomeres), after amanitin treatment (two biological replicates consisting of two specimens) (more than fivefold, $P < 0.01$), and parthenotes (one sample consisting of four specimens). Data analysis for downregulated transcripts was done accordingly: 829 genes with transcript levels of 20% or less of those found in the oocyte were identified ($P < 0.01$). Among those 829 transcripts, we determined the number of transcripts that were also downregulated after genome exchange and after amanitin treatment ($P < 0.01$). Gene expression analysis of soPS cells was done with normalization to average and subtraction of background.

All array data and additional details on samples and analysis are available on GEO under accession number GSE28024.

Cell line analysis. For karyotype and STR analysis, live cultures were shipped to Cell Line Genetics (WI). RNA and DNA were isolated from cultures at passage 7 to 11, using QIAamp DNA Mini Kit for SNP capture. Padlock probes (27,000) were prepared to capture expressed SNPs from genomic DNA and cDNAs. The captured SNPs were quantified using single-molecule DNA sequencing according to ref. 30. Immunohistochemistry was done using primary antibodies recognizing Tra1-60 (MAB4360, Millipore), Tra1-81 (MAB5381, Millipore), SSEA-4 (MAB1435, R&D), SSEA-3 (MAB1434, R&D), Nanog (AF1997, R&D), Oct-4 (09-0023, Stemgent), Sox2 (09-0024, Stemgent), MF20 (DSHB), AFP (Dako), beta III tubulin (Sigma T2200) at dilutions of 1:500 to 1:1,000. Secondary antibodies were conjugated with Alexa Fluor (Invitrogen). Alkaline phosphatase staining was done using an alkaline phosphatase substrate kit (Vector Laboratories). Teratomas were generated by subcutaneous injection into NSG mice (Jackson laboratories) and harvested after 10–15 weeks. Animal experimentation was approved by the Columbia IACUC. For Affymetrix SNP chip analysis, gDNA was processed according to the GeneChip mapping 500K assay manual, and hybridized to a 250K Nsp Array according to the manufacturer's instructions. SNP data analysis was done using Affymetrix Genotyping Console. Bisulphite conversion of DNA was done using the EpiTect Bisulfite Kit according to the manufacturer's instructions. For bisulphite sequencing, PCR products were cloned into Topo TA vector (Invitrogen) and sequenced using M13R primer.

Primers used in this study were as follows: Primer sequence 5' to 3': ATTTGTTTTTTGGGTAGTTAAAGGT and CCTAACTCCCCTTCAAATCTATT, bisulphite sequencing of Oct4 (ref. 36); TGG TTA GGT TGG TTT TAA ATT TTT G and AAC CCA CCC TTA TAA ATT CTC AAT TA, bisulphite sequencing of Nanog³⁷; GGAAAGAAATTTTATAGGTAGGATAGT and AAACCTAAACCTCCTAACTAACTCTAA, bisulphite sequencing of PEG3 (ref. 38); GGGGTGAGTGCCTCAATAAG and TTTGGTCTCCAGTTTCAGG, sequencing of rs7221396; CAGTTTACCCCTTACCTTCA and ACACATGTTGCCACCAGAGA, sequencing of rs11652263; GGGTTAGAAGCTCCTGCAAA and CTCTGGTCTGTCAACCATCA, sequencing of rs2286336; CAGTTTGACAGACATGAAATCC and CTGCTTTTCTGCTGCAATTGT, sequencing of rs10521202; CACCATTAGCACCCAAAGCT and TGATTTACGGAGGATGGTG, sequencing of mitochondrial hypervariable region I.

Images and settings. GFP fluorescence and bright field images were acquired with a Nikon TE 2000-U microscope equipped with a Nikon Digital Sight DS-Qi1Mc camera and NIS elements AR imaging program. GFP fluorescence was acquired with 1-s exposure time, no gain for all images. Bright field images were contrast-adjusted (equally across the entire image) in Adobe Photoshop. Histology analysis was done using an Olympus IX-71 microscope equipped with a U-TV0.5XC-3 colour camera. Immunofluorescence was done using Olympus DP30BW camera and Olympus imaging acquisition software. All fluorescent images are pseudo colours. Figures were assembled in Adobe Freehand MX.

33. Chen, A. E. *et al.* Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell Stem Cell* **4**, 103–106 (2009).

34. Cowan, C. A. *et al.* Derivation of embryonic stem-cell lines from human blastocysts. *N. Engl. J. Med.* **350**, 1353–1356 (2004).

35. Dimos, J. T. *et al.* Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321**, 1218–1221 (2008).
36. Freberg, C. T., Dahl, J. A., Timoskainen, S. & Collas, P. Epigenetic reprogramming of *OCT4* and *NANOG* regulatory regions by embryonal carcinoma cell extract. *Mol. Biol. Cell* **18**, 1543–1553 (2007).
37. Imamura, M. *et al.* Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *BMC Dev. Biol.* **6**, 34 (2006).
38. Pick, M. *et al.* Clone- and gene-specific aberrations of parental imprinting in human induced pluripotent stem cells. *Stem Cells* **27**, 2686–2690 (2009).

Complement factor H binds malondialdehyde epitopes and protects from oxidative stress

David Weismann^{1,2}, Karsten Hartvigsen^{1,2,3}, Nadine Lauer⁴, Keiryn L. Bennett¹, Hendrik P. N. Scholl⁵, Peter Charbel Issa⁶, Marisol Cano⁵, Hubert Brandstätter^{2,7}, Sotirios Tsimikas³, Christine Skerka⁴, Giulio Superti-Furga¹, James T. Handa⁵, Peter F. Zipfel⁴, Joseph L. Witztum³ & Christoph J. Binder^{1,2,3}

Oxidative stress and enhanced lipid peroxidation are linked to many chronic inflammatory diseases, including age-related macular degeneration (AMD). AMD is the leading cause of blindness in Western societies, but its aetiology remains largely unknown. Malondialdehyde (MDA) is a common lipid peroxidation product that accumulates in many pathophysiological processes, including AMD. Here we identify complement factor H (CFH) as a major MDA-binding protein that can block both the uptake of MDA-modified proteins by macrophages and MDA-induced proinflammatory effects *in vivo* in mice. The CFH polymorphism H402, which is strongly associated with AMD, markedly reduces the ability of CFH to bind MDA, indicating a causal link to disease aetiology. Our findings provide important mechanistic insights into innate immune responses to oxidative stress, which may be exploited in the prevention of and therapy for AMD and other chronic inflammatory diseases.

Increased oxidative stress has been implicated in the pathogenesis of many different diseases¹. As a consequence of oxidative stress, proteins, lipids and DNA can be damaged, often resulting in structural changes. For example, when membrane phospholipids undergo lipid peroxidation, MDA and other reactive decomposition products are generated². These can in turn modify endogenous molecules, generating novel oxidation-specific epitopes (OSEs), which are also present on the surface of apoptotic cells and blebs released from them³. Many of these OSEs are recognized as danger signals by innate immune receptors⁴. Elucidating the molecular mechanisms by which oxidative damage challenges the immune system would pave the road for new diagnostic and therapeutic approaches in several pathologies.

MDA and its condensation products are reliable markers for oxidative stress and have been associated with many disorders, including atherosclerosis^{1,4} and AMD, a degenerative disease affecting the retina that leads to irreversible vision loss^{5,6}. AMD is the most common cause of blindness in the elderly in Western societies⁷. A hallmark of developing AMD is the accumulation of extracellular deposits, termed drusen, which have been shown to contain MDA⁸. MDA-modified proteins are known to induce inflammatory responses and are recognized by innate immunity^{9–11}. We recently demonstrated that OSEs in general are a major target of innate natural antibodies both in mice and humans and that ~15% of all immunoglobulin M (IgM) natural antibodies bound MDA-type adducts, suggesting a great need to defend against this specific modification¹². However, the abundance of MDA and the danger associated with it suggests that additional, evolutionary conserved innate defence mechanisms exist.

CFH binds MDA modifications

We used an unbiased proteomic approach to identify plasma proteins binding to MDA modifications. Because normal plasma contains high

titres of MDA-specific natural antibodies¹², we purified MDA-binding proteins from plasma of atherosclerotic *Rag*^{−/−}*Ldlr*^{−/−} mice that lack immunoglobulins. Pooled plasma was incubated with beads coupled to either malondialdehyde-acetaldehyde (MAA)-modified or unmodified polylysine, respectively. MAA is an advanced MDA-lysine adduct the structure of which is shown in Supplementary Fig. 1 (see also ref. 13). Bound proteins were eluted and identified by mass spectrometry. As many as 45 unique peptides were found exclusively in MAA-polylysine pull-downs, of which >55% could be attributed to CFH (Supplementary Fig. 2 and Supplementary Table 1). CFH is a major regulator of the complement system and protects host tissues from complement-mediated damage¹⁴. Immunoblot analysis revealed the presence of CFH on MAA-coated beads but not on control beads (Fig. 1a). This finding was confirmed using human plasma (Fig. 1b). Interestingly, the anti-CFH antibody also detected lower molecular weight bands, which may represent CFH-related proteins (CFHRs) that share high sequence homology with CFH.

Using enzyme-linked immunosorbent assay (ELISA), we demonstrated that CFH bound to MDA directly and independently of the protein carrying the adducts. Purified CFH bound in a calcium-independent manner to both MAA-modified low density lipoprotein (MAA-LDL) and MAA-modified bovine serum albumin (MAA-BSA), but not to unmodified proteins (Fig. 1c and Supplementary Fig. 3). Moreover, we tested binding of CFH to the oxidation-specific modifications phosphocholine-BSA (PC-BSA), which is bound by C reactive protein (CRP), as well as carboxyethylpyrrole-BSA (CEP-BSA) and 4-hydroxynonenal-BSA (4-HNE-BSA). None of these modifications were bound by CFH (Fig. 1d and Supplementary Fig. 4A, B). CRP and C3 were also detected in the MAA-polylysine pull-downs (Supplementary Table 1), but neither of them bound to coated MAA-BSA (Fig. 1d and Supplementary Fig. 4C).

¹Center for Molecular Medicine (CeMM) of the Austrian Academy of Sciences, 1090 Vienna, Austria. ²Department of Laboratory Medicine, Medical University of Vienna, 1090 Vienna, Austria. ³Department of Medicine, University of California at San Diego, La Jolla, California 92093, USA. ⁴Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute and Friedrich Schiller University, 07745 Jena, Germany. ⁵Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA. ⁶Nuffield Laboratory of Ophthalmology, University of Oxford, Oxford OX3 9DU, UK. ⁷Octapharma PPGmbH, Research & Development, 1100 Vienna, Austria.

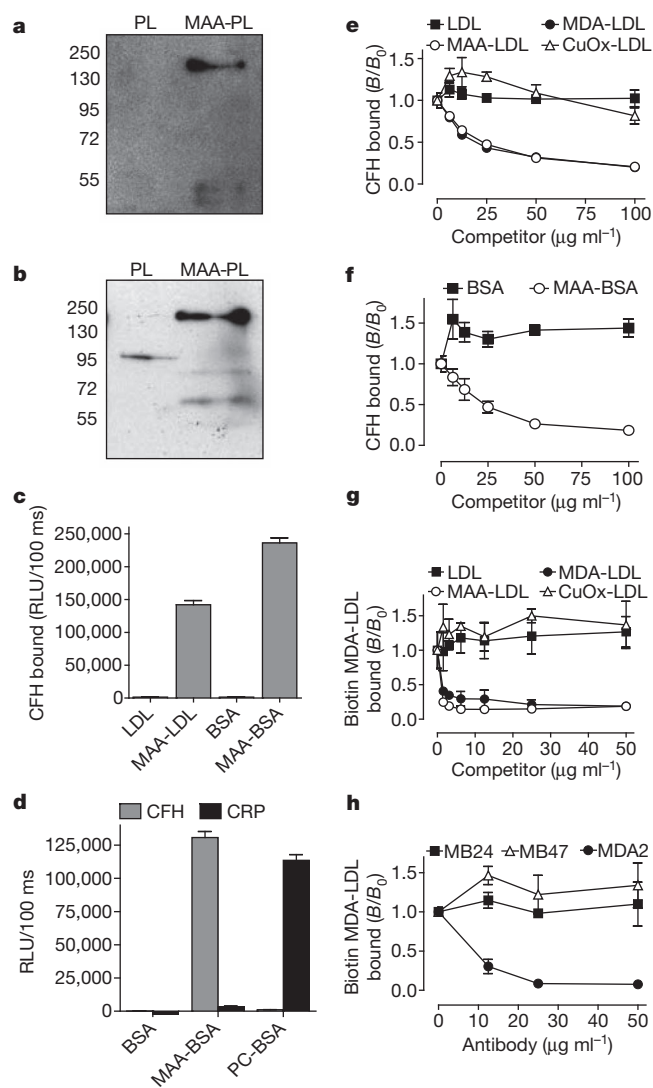


Figure 1 | CFH specifically binds to MDA modifications. **a, b**, Immunoblot for CFH (molecular weight: 150 kDa) using eluates from either polylysine (PL) or MAA-polylysine (MAA-PL) beads incubated with *Ldlr*^{-/-}*Rag*^{-/-} mouse plasma (**a**) or human plasma (**b**). **c**, ELISA for binding of purified CFH (5 $\mu\text{g ml}^{-1}$) to coated native LDL, MAA-LDL, BSA and MAA-BSA. Values are mean \pm s.d. relative light units (RLU) per 100 ms of triplicate determinations. **d**, ELISA for binding of purified CFH or CRP (5 $\mu\text{g ml}^{-1}$) to coated BSA, MAA-BSA and PC-BSA. Values are mean \pm s.d. RLU per 100 ms of triplicate determinations. **e–h**, Competition immunoassays. **e, f**, Binding of purified CFH (**e**) or binding of plasma CFH (**f**) to coated MAA-BSA in the presence of increasing concentrations of LDL, MDA-LDL, MAA-LDL and Cu²⁺-oxidized (CuOx)-LDL, or BSA and MAA-BSA. **g**, Binding of biotinylated MDA-LDL to coated CFH in the presence of increasing concentrations of LDL, MDA-LDL, MAA-LDL and CuOx-LDL. **h**, Binding of biotinylated MDA-LDL to coated CFH in the presence of increasing concentrations of monoclonal antibodies specific for ApoB100 (MB24 and MB47) or MDA (MDA2). Data are expressed as a ratio of binding in the presence of competitor divided by the binding in the absence of competitor (B/B_0) and represent the mean \pm s.d. of triplicate determinations. As an estimate for the affinity, the dissociation constants K_d were calculated as $6.4 \times 10^{-8} \text{ mol l}^{-1}$ for the binding of CFH to coated MAA-BSA and $1.6 \times 10^{-9} \text{ mol l}^{-1}$ for the binding of MAA-BSA to coated CFH.

To characterize the specificity of the binding of CFH to MDA, we performed competition assays. Only MDA- and MAA-modified LDL competed in a concentration-dependent manner for the binding of CFH to coated MAA-BSA. Neither native LDL nor the negatively charged Cu²⁺-oxidized LDL showed inhibition, thereby excluding non-specific interactions mediated by charge effects (Fig. 1e). To demonstrate a dose-dependent interaction in plasma, the binding of

CFH to coated MAA-BSA was tested in different plasma dilutions (Supplementary Fig. 5). Consistent with the notion that CFH is a major MDA-binding protein in plasma, binding of CFH to coated MAA-BSA was competed by soluble MAA-BSA with similar efficiency in whole plasma (Fig. 1f). In a reciprocal experiment, binding of biotinylated MDA-LDL to immobilized purified CFH was fully competed by either MDA- or MAA-modified LDL, even at very low competitor concentrations (Fig. 1g). In the same assay, the MDA-lysine-specific monoclonal antibody MDA2 fully inhibited binding of MDA-LDL to CFH. In contrast, the apoB-100-specific monoclonal antibodies MB47 and MB24, which bind apoB-100 of MDA-LDL (Supplementary Fig. 6), did not inhibit this interaction (Fig. 1h). Using surface plasmon resonance, we observed a concentration-dependent binding of CFH to coated MAA-BSA (Supplementary Fig. 7). Taken together, these findings prove that CFH binds specifically to MDA modifications.

The CFH H402 variant has impaired MDA binding

To map the binding site for MDA on CFH, we performed binding studies using recombinantly expressed CFH fragments (Fig. 2a). CFH is composed of 20 globular short consensus repeats (SCRs)¹⁵. Only fragments containing either SCR7 or SCR20 bound to coated MDA (Fig. 2a). Reciprocally, soluble MDA-LDL only bound to immobilized fragments containing either SCR7 or SCR20, respectively (Supplementary Fig. 8). Importantly, these domains have also been identified as clustering points of various disease-related mutations¹⁵.

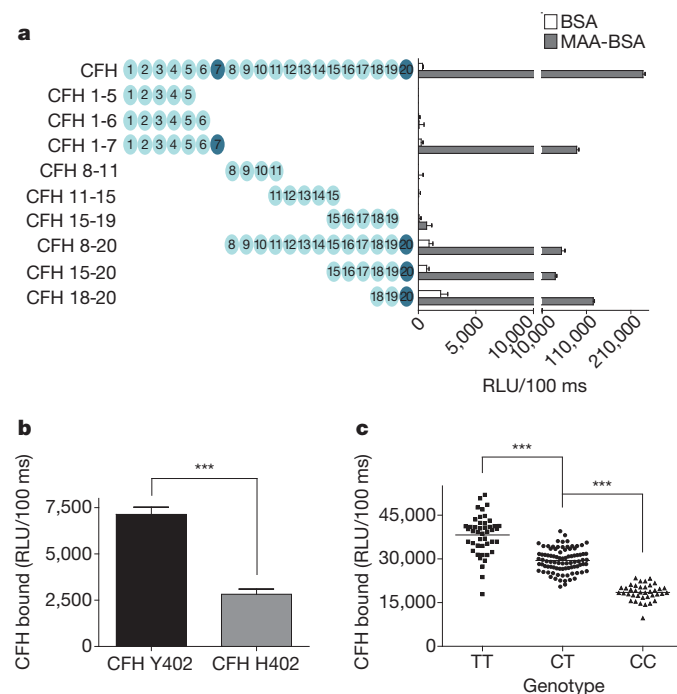


Figure 2 | The SCR7 domain of CFH is critical for MDA binding. **a**, ELISA for binding of CFH and recombinantly expressed CFH fragments to coated BSA (white bars) or MAA-BSA (black bars). The length of CFH fragments is indicated by schematic representations with each circle depicting one SCR. Values are mean \pm s.d. RLU per 100 ms of triplicate determinations. **b**, ELISA for binding of purified CFH variant Y402 and CFH variant H402 (both at 1 $\mu\text{g ml}^{-1}$) to coated MAA-BSA. Values are mean \pm s.d. RLU per 100 ms of triplicate determinations. **c**, ELISA for binding of plasma CFH to coated MDA-LDL in plasma of subjects homozygous for the H402 risk allele (CC, $n = 38$), heterozygous for the H402 risk allele (CT, $n = 88$) or homozygous for the Y402 allele (TT, $n = 45$). The association of rs1061170 with CFH binding to MDA was calculated with $P = 1.29 \times 10^{-40}$ using an additive model. Symbols represent individual subject samples with horizontal bars indicating the mean of each group. Values are mean \pm s.d. RLU per 100 ms of triplicate determinations (***) $P < 0.001$.

One of the most widely studied single nucleotide polymorphisms (SNPs) in *CFH* is the prevalent rs1061170 SNP, which causes an amino acid switch on position 402 (Y→H) in SCR7. To determine the effect of the H402 substitution, we purified CFH from plasma of homozygous individuals expressing either CFH Y402 or CFH H402, respectively, and tested the binding to MDA. Compared to the common Y402 variant, the CFH variant H402 exhibited significantly impaired binding to MAA-BSA (Fig. 2b). The H402 variant has been associated with a significant risk for the development of AMD^{16–19}. Therefore, we analysed the binding of CFH to coated MDA-LDL in plasma samples of AMD patients with the respective genotypes. Compared to the extent of CFH binding to MDA-LDL using plasma of individuals homozygous for the protective allele, binding in plasma of heterozygous subjects was reduced by 23% ($P < 0.001$), and by 52% ($P < 0.001$) in plasma of subjects homozygous for the H402 risk allele (Fig. 2c), irrespective of the total plasma CFH levels (Supplementary Fig. 9A). Moreover, plasma levels of MDA-specific IgG and IgM antibodies were similar in all groups (Supplementary Fig. 9B, C).

The genetic deletion of *CFHR1* and *CFHR3* has been reported to protect from AMD and could influence CFH binding to MDA²⁰. Less than 25% of individuals in this study carried deletions at these loci and their removal from our analysis did not alter the significance of the association of rs1061170 with MDA binding (Supplementary Fig. 9D). Taken together, the impaired ability of the risk variant to bind MDA suggested an important role for this interaction in AMD pathogenesis.

CFH binds cellular debris via MDA epitopes

Owing to constant light exposure, the retina provides an environment that facilitates lipid peroxidation⁷. We detected MDA epitopes by immunohistochemistry in the eyes of subjects with and without AMD. MDA epitopes were detectable throughout the choroid and Bruch's membrane (Fig. 3a, d). In eyes without AMD, labelling for MDA was stronger in the outer than inner Bruch's membrane (Fig. 3a). In eyes with AMD, MDA staining was seen diffusely throughout Bruch's membrane (Fig. 3d). Staining for CFH followed a similar pattern (Fig. 3b, e). In addition, strong CFH labelling was seen in the retinal pigment epithelium (RPE) and choriocapillaris basement membranes. Moreover, the presence of C3d, a cleavage product of iC3b, indicated co-factor activity at the same sites (Fig. 3c, f). We further demonstrated by confocal microscopy the presence of MDA epitopes on the surface of *in vitro*-generated necrotic RPE cells, a major cell type affected in AMD. Moreover, CFH colocalized with MDA epitopes, suggesting that MDA mediates recognition of dying cells by CFH (Fig. 3j). To demonstrate this directly, we used flow cytometry to assess the binding of CFH to apoptotic blebs from Jurkat T cells in the presence of MAA-BSA as competitor. Consistent with the presence of MDA epitopes on only a subgroup of apoptotic blebs, we found that CFH bound between 5–45% of apoptotic blebs (Supplementary Fig. 10). Importantly, MAA-BSA competed for this binding by more than 60%, whereas unmodified BSA did not (Fig. 3k). Thus, MDA adducts present in several retinal compartments and on the surface of necrotic RPE cells represent *in vivo* ligands for CFH. In addition, we predicted that CFH might also bind to MDA adducts in other tissues and confirmed this in atherosclerotic lesions (Supplementary Fig. 11).

CFH inactivates complement on MDA-bearing surfaces

An important regulatory activity of CFH lies with its capacity to act as co-factor for serine protease factor I, thereby promoting the degradation of C3b into inactive iC3b fragments. Deposition of iC3b on apoptotic cells increases their clearance in an anti-inflammatory manner^{21,22}. We therefore tested whether CFH induces iC3b generation when bound to MDA. Indeed, CFH promoted the formation of iC3b in a dose- and time-dependent manner when bound to coated MAA-BSA (Supplementary Fig. 12). When comparing the co-factor activity of the 402 variants on MDA-decorated surfaces, we discovered

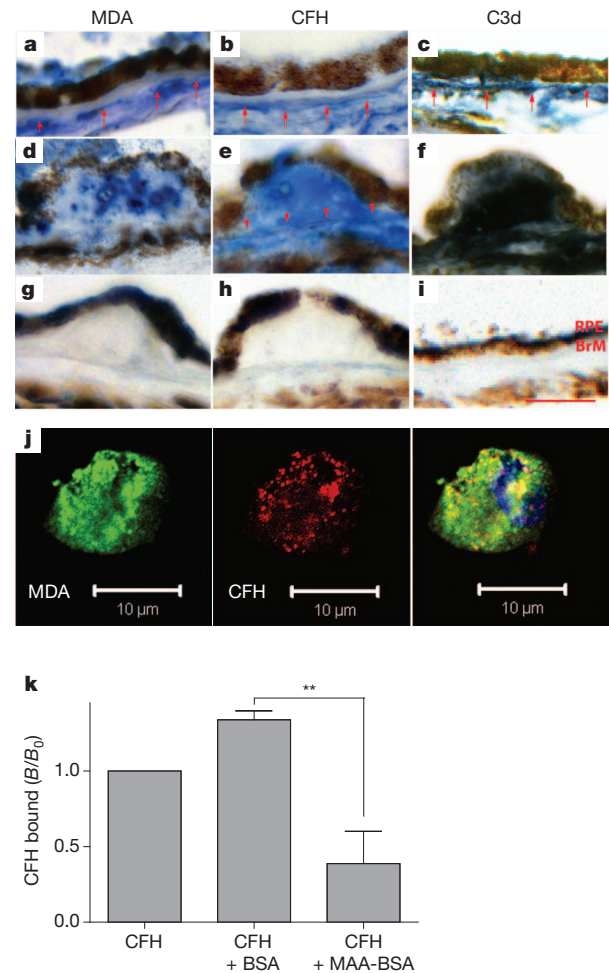


Figure 3 | CFH binds to MDA epitopes present in AMD lesions, on necrotic cells and apoptotic blebs. a–i, Immunohistochemistry of MDA (left), CFH (middle) and C3d (right) localization in human maculas. a–c, Eye of a 72-year-old subject heterozygous for the H402 SNP without AMD. d–f, Eye of a 93-year-old subject homozygous for the H402 SNP with AMD. g–i, IgG control immunostains. Red arrows indicate positive labelling of choriocapillaris basement membrane and arrowheads indicate labelling of Bruch's membrane (BrM). Scale bar, 25 μm. Sections are representative for 7 donors (5 AMD, 2 controls). j, Confocal immunofluorescent photograph of necrotic RPE cells stained with the MDA-specific IgM natural antibody EO14 (green) and CFH (red), respectively. The right panel shows a merged picture indicating co-localization of CFH binding with the presence of MDA epitopes (yellow). k, Competition assay for the binding of CFH to apoptotic blebs from Jurkat T cells either alone or in the presence of BSA or MAA-BSA assessed by flow cytometry. Values are expressed as mean \pm s.e.m. CFH binding (B/B_0) based on mean fluorescence intensities of four independent experiments (** $P < 0.01$).

a strong functional difference in that impaired MDA-binding of the risk variant resulted in severely reduced factor-I-mediated C3 cleavage (Fig. 4a). This activity of CFH may represent an important protective mechanism in conditions in which MDA is continuously generated, for example, on the surface of dying cells.

Importantly, other members of the CFH family such as *CFHR1*/*CFHR3* show homology with the carboxy terminus of CFH and therefore contain a potential MDA-binding site without possessing co-factor activity. Deletions of *CFHR1/3* have been reported to be protective in AMD²⁰, suggesting a negative role of these proteins in this pathology. To demonstrate the potential capacity of MDA-binding CFHR to inhibit the beneficial co-factor activity of CFH, we tested whether C-terminal CFH fragments could compete for the co-factor activity by binding to MDA. Indeed, the MDA-binding fragment SCR18–20 prevented CFH from inducing iC3b generation, whereas a non-binding fragment containing SCR15–19 had no effect

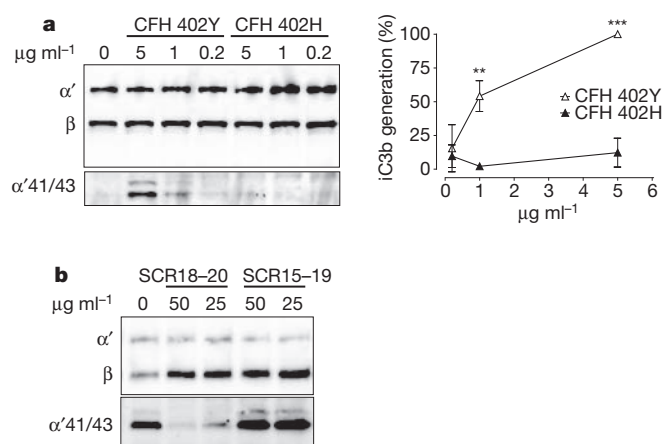


Figure 4 | CFH inactivates complement on MDA-bearing surfaces.

a, Immunoblot of C3b cleavage products induced by three concentrations of CFH variants bound to coated MAA-BSA. $\alpha'41/43$ indicates iC3b cleavage products of the C3b α -chain. β indicates the C3b β -chain that remains uncleaved and served as a loading control. $\alpha'41/43$ was densitometrically quantified, and data are presented as percentage of iC3b generation achieved with $5 \mu\text{g ml}^{-1}$ CFH Y402 (= 100%). Shown is the mean \pm s.e.m. of three independent experiments. **b**, Immunoblot of C3b degradation products induced by CFH bound to coated MAA-BSA in the presence of either CFH 18-20 or CFH 15-19.

(Fig. 4b). These data point towards a complex regulation of complement activation on MDA-decorated surfaces.

CFH neutralizes proinflammatory effects of MDA

The inflammatory process in AMD lesions has been suggested to be propagated by the secretion of cytokines including IL-8 (ref. 23). Stimulation of RPE cells (ARPE-19) with MAA-BSA induced the expression of IL-8 and caused an antioxidant response as indicated by upregulation of NAD(P)H dehydrogenase and hemoxygenase-1 (Fig. 5a). We then tested the effect of CFH on MAA-LDL binding to macrophages, another cell type involved in AMD pathogenesis²⁴. In a cell-based ELISA, CFH inhibited binding of MAA-LDL to thioglycollate-elicited macrophages in a dose-dependent manner (Fig. 5b). This indicates that CFH binds the same epitope on MAA-LDL that is necessary for its recognition by macrophages. Similar to ARPE-19 cells, monocytic THP-1 cells exhibited a robust expression of IL-8 following MAA-BSA stimulation. In addition, MAA-BSA induced the expression of TNF- α and IL-1 β , but not IL-12 β (Supplementary Fig. 13). Importantly, MAA-BSA-induced IL-8 secretion was inhibited by physiological concentrations of CFH in a dose-dependent manner (Fig. 5c). In contrast, CFH had no effect on IL-8 production induced by phorbol myristate acetate (Supplementary Fig. 14).

To evaluate the importance of this interaction *in vivo*, we examined the effect of MAA adducts in a murine model. First we validated that MAA-BSA could induce secretion of KC, the mouse orthologue to human IL-8, in murine macrophages (Supplementary Fig. 15). To test the proinflammatory effect of MAA-BSA as well as the scavenging capacity of CFH in an AMD-relevant site, we performed intravitreal microinjections of MAA-BSA with or without CFH. After six hours mice were killed, RPE/choroid was isolated from each eye after enucleation, and RNA was extracted. The purity of the preparation was confirmed by the expression of the RPE-specific gene *Rpe65* and the lack of rhodopsin (*Rho*) expression as a marker for the neurosensory retina in all samples (Fig. 5d). MAA-BSA injection led to a sevenfold upregulation of KC expression in these RPE preparations, whereas BSA injection had no effect. Importantly, addition of CFH completely inhibited the effect induced by MAA-BSA (Fig. 5e). Thus, MAA adducts promote inflammatory responses in different cell types involved in AMD *in vitro* and in the eye *in vivo*, and CFH specifically neutralizes this property.

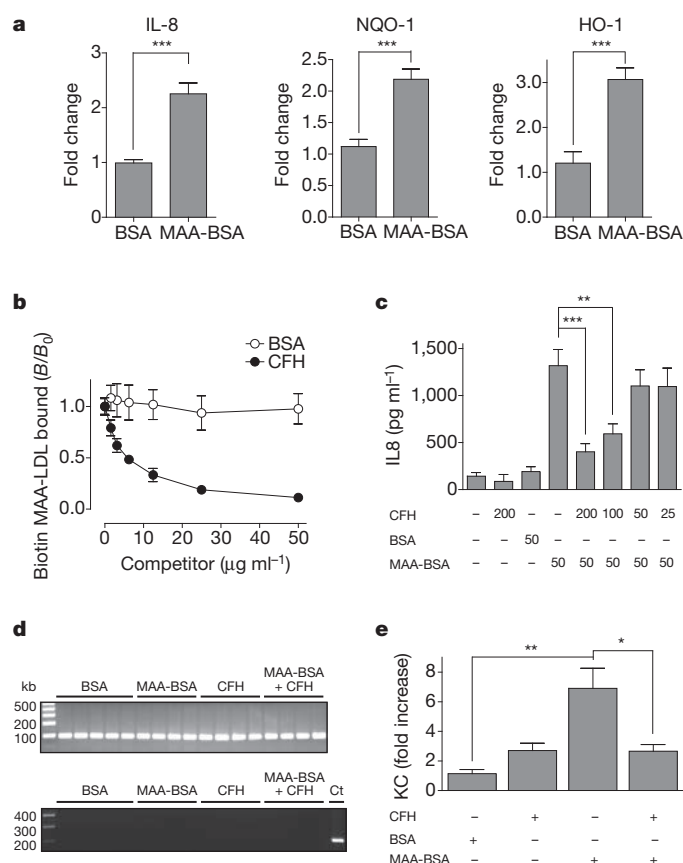


Figure 5 | CFH neutralizes proinflammatory effects of MDA. **a**, Expression of indicated genes in ARPE-19 cells stimulated for 24 h with $50 \mu\text{g ml}^{-1}$ BSA compared to MAA-BSA as determined by quantitative RT-PCR. Data represent the mean \pm s.e.m. of three independent experiments. **b**, Cell-based ELISA for the binding of biotinylated MAA-LDL to thioglycollate-elicited macrophages in the presence of BSA or CFH. Data are expressed as B/B_0 and represent mean \pm s.d. of triplicate determinations. **c**, Secretion of IL-8 by THP-1 cells stimulated for 12 h with BSA or MAA-BSA in the absence or presence of CFH. Numbers below indicate concentrations of CFH, BSA and MAA-BSA in $\mu\text{g ml}^{-1}$. Error bars represent mean \pm s.e.m. of three independent experiments. **d**, **e**, Intravitreal injection of BSA, MAA-BSA and/or CFH in mice ($n = 4-5$ per group). Six hours after injection, RPE/choroid was isolated. **d**, RT-PCR for *Rpe65* and *Rho* in RPE/choroid fractions. cDNA isolated from neurosensory retina was used as a control (Ct). **e**, Expression of KC in RPE/choroid as assessed by quantitative RT-PCR. Error bars represent mean \pm s.e.m. expression normalized to the BSA-injected group (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

Discussion

We report the identification of CFH as a hitherto unrecognized innate defence protein against MDA, which is a ubiquitously generated proinflammatory product of lipid peroxidation^{1,9,10}. Our discovery of CFH as a major MDA-binding protein demonstrates that innate immunity has a pivotal role in providing homeostatic responses against endogenous oxidation-specific danger-associated molecular patterns⁴. This parallels the innate immune response to another OSE, the PC headgroup of oxidized phosphatidylcholine, mediated by the macrophage scavenger receptors CD36 and SR-B1, the murine IgM natural antibody EO6/T15 and the acute phase reactant CRP^{4,25}. In an analogous manner, MDA is recognized by macrophage scavenger receptor SR-A¹¹, several germline IgM natural antibodies¹² and—as we now demonstrate—by CFH.

CFH is one of the most abundant plasma proteins ($\sim 100-700 \mu\text{g ml}^{-1}$) and the major regulator of complement activation¹⁴. It mediates anti-inflammatory housekeeping functions by protecting self cells from complement activation¹⁴, which is especially important

for dying cells that lose other surface-associated complement regulators^{26,27}. A number of potential ligands for CFH on host cells have been studied, including glycosaminoglycans²⁸, as well as annexin A2, DNA and histones on apoptotic cells²⁹. We now identify MDA as a major ligand for CFH on apoptotic/necrotic cells and show that MDA epitopes provide a surface for CFH to allow local generation of anti-inflammatory iC3b fragments. This becomes relevant in situations when large amounts of cellular debris are generated. Of note, necrotic and, under certain conditions, apoptotic cells are proinflammatory per se^{30,31}. In this regard, the interaction of CFH with MDA-modified cellular compounds is also important because CFH limits MDA-induced IL-8 secretion. This provides an explanation for the ability of CFH to reduce endothelial IL-8 secretion in response to apoptotic blebs³². Thus, MDA epitopes are responsible for the recruitment of CFH to the surface of apoptotic cells, where it neutralizes their proinflammatory properties and halts complement activation.

We demonstrate that SCR7 and SCR20 mediate the binding of CFH to MDA. These two domains are clustering sites for mutations associated with AMD, but also other diseases¹⁴. The most prominent example is the H402 exchange in SCR7, which has a frequency of 35%, and may be responsible for over half of all AMD cases^{33,34}. However, direct evidence for functional consequences of this polymorphism remained elusive. Here we show that the H402 variant exhibits severely impaired binding to MDA in a gene-dosage-dependent manner, which correlates well with the H402-associated risk for developing AMD. The H402 variant has been suggested to favour local complement activation as a result of reduced binding to glycosaminoglycans in the eye³⁵. However, in contrast to glycosaminoglycans, MDA is enriched in the membranes of dying cells, which are continuously generated in the retina and need to be efficiently removed³⁶. By demonstrating that the H402 variant has a reduced capacity to generate anti-inflammatory iC3b fragments on MDA-bearing surfaces, we provide a functional explanation for its strong disease association. It remains to be seen whether other genetic variations of CFH family members also affect MDA binding and thereby contribute to disease pathogenesis.

Consistent with an earlier report, we found MDA epitopes throughout the choroid and Bruch's membrane including drusen of AMD lesions⁸. However, even under physiological conditions, oxidized phospholipids are formed as a result of photic stimulation of retinal photoreceptors and subsequently scavenged by several processes including clearance via CD36 (ref. 37). As one of the major degradation products of peroxidized phospholipids, MDA is continuously generated. Several lipid peroxidation products, including MDA, can cause RPE damage³⁸. Therefore, physiological housekeeping mechanisms are critically needed to prevent their accumulation and adverse reactions mediated by them. Our immunohistochemical results support a role for CFH, because CFH is found in the same locations as MDA in eyes with and without AMD. We show that MDA adducts, similar to what has been shown following the ingestion of oxidized photoreceptors²³, have the capacity to induce IL-8 secretion in RPE, which can be blocked by CFH. Increased IL-8 expression correlates with higher incidence of AMD, underlining its important pathogenic role³⁹. Therefore, neutralization of MDA adducts by CFH has the potential to limit several pathogenic events in AMD.

Undoubtedly, there are multiple defences against the ubiquitous MDA adducts. The described homeostatic response may be particularly limiting in the eye, as opposed to other sites where MDA adducts accumulate such as the vascular wall⁴⁰. Future studies need to evaluate the contribution of this newly found interaction in the pathogenesis of other inflammatory diseases. The findings described here may lead to novel approaches exploiting endogenous defence mechanisms for the prevention and therapy of chronic inflammation in general.

METHODS SUMMARY

Subjects and clinical diagnosis. The patient cohort used in this study has been described⁴¹.

Protein modifications. The MAA modifications of LDL, BSA or polylysine were performed as described¹³.

Intravitreal injection. Intravitreal injections of BSA, MAA-BSA and/or CFH were performed in male and female C57BL/6 mice. Six hours later, the RPE/choroid was isolated and expression of target genes assessed by quantitative RT-PCR.

Bead coupling and pull-down procedure. Mouse or human plasma was incubated with polylysine or MAA-polylysine beads. Bound proteins were analysed by LC-MS/MS. The interaction was verified by immunoblotting and characterized by ELISA and Biacore.

Flow cytometry and immunohistochemistry. The presence of MDA epitopes and CFH was visualized by flow cytometry on apoptotic Jurkat T-cell microparticles and by immunohistochemistry for histological specimens.

Co-factor assay. CFH bound to coated MAA-BSA was incubated with C3b and factor I and the generation of iC3b fragments was visualized by immunoblotting.

Cell culture. Following stimulation with BSA or MAA-BSA in the presence or absence of CFH, gene expression was determined by quantitative RT-PCR and/or IL-8/KC secretion was quantified by ELISA. Binding of biotinylated MAA-LDL to peritoneal macrophages was assessed in the presence or absence of competitors as described⁴².

Statistical analysis. Data are presented as mean \pm s.d. or mean \pm s.e.m. where indicated. Results were analysed by one-way analysis of variance and Student's unpaired *t*-test.

Received 25 January; accepted 12 August 2011.

- Chou, M. Y. *et al.* Oxidation-specific epitopes are important targets of innate immunity. *J. Intern. Med.* **263**, 479–488 (2008).
- Esterbauer, H., Schaur, R. J. & Zollner, H. Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. *Free Radic. Biol. Med.* **11**, 81–128 (1991).
- Chang, M. K. *et al.* Monoclonal antibodies against oxidized low-density lipoprotein bind to apoptotic cells and inhibit their phagocytosis by elicited macrophages: evidence that oxidation-specific epitopes mediate macrophage recognition. *Proc. Natl Acad. Sci. USA* **96**, 6353–6358 (1999).
- Miller, Y. I. *et al.* Oxidation-specific epitopes are danger-associated molecular patterns recognized by pattern recognition receptors of innate immunity. *Circ. Res.* **108**, 235–248 (2011).
- Hollyfield, J. G. *et al.* Oxidative damage-induced inflammation initiates age-related macular degeneration. *Nature Med.* **14**, 194–198 (2008).
- Suzuki, M. *et al.* Oxidized phospholipids in the macula increase with age and in eyes with age-related macular degeneration. *Mol. Vis.* **13**, 772–778 (2007).
- Jager, R. D., Mieler, W. F. & Miller, J. W. Age-related macular degeneration. *N. Engl. J. Med.* **358**, 2606–2617 (2008).
- Schutt, F., Bergmann, M., Holz, F. G. & Kopitz, J. Proteins modified by malondialdehyde, 4-hydroxynonenal, or advanced glycation end products in lipofuscin of human retinal pigment epithelium. *Invest. Ophthalmol. Vis. Sci.* **44**, 3663–3668 (2003).
- Thiele, G. M. *et al.* Malondialdehyde-acetaldehyde (MAA) modified proteins induce pro-inflammatory and pro-fibrotic responses by liver endothelial cells. *Comp. Hepatol.* **3** (suppl. 1), S25 (2004).
- Shanmugam, N. *et al.* Proinflammatory effects of advanced lipoxidation end products in monocytes. *Diabetes* **57**, 879–888 (2008).
- Shechter, I. *et al.* The metabolism of native and malondialdehyde-altered low density lipoproteins by human monocyte-macrophages. *J. Lipid Res.* **22**, 63–71 (1981).
- Chou, M. Y. *et al.* Oxidation-specific epitopes are dominant targets of innate natural antibodies in mice and humans. *J. Clin. Invest.* **119**, 1335–1349 (2009).
- Xu, D. *et al.* Epitope characterization of malondialdehyde-acetaldehyde adducts using an enzyme-linked immunosorbent assay. *Chem. Res. Toxicol.* **10**, 978–986 (1997).
- Zipfel, P. F. & Skerka, C. Complement regulators and inhibitory proteins. *Nature Rev. Immunol.* **9**, 729–740 (2009).
- Józsi, M. & Zipfel, P. F. Factor H family proteins and human diseases. *Trends Immunol.* **29**, 380–387 (2008).
- Haines, J. L. *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308**, 419–421 (2005).
- Klein, R. J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389 (2005).
- Edwards, A. O. *et al.* Complement factor H polymorphism and age-related macular degeneration. *Science* **308**, 421–424 (2005).
- Hageman, G. S. *et al.* A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc. Natl Acad. Sci. USA* **102**, 7227–7232 (2005).
- Hughes, A. E. *et al.* A common CFH haplotype, with deletion of CFHR1 and CFHR3, is associated with lower risk of age-related macular degeneration. *Nature Genet.* **38**, 1173–1177 (2006).
- Takizawa, F., Tsuji, S. & Nagasawa, S. Enhancement of macrophage phagocytosis upon iC3b deposition on apoptotic cells. *FEBS Lett.* **397**, 269–272 (1996).
- Amarilyo, G. *et al.* iC3b-opsonized apoptotic cells mediate a distinct anti-inflammatory response and transcriptional NF- κ B-dependent blockade. *Eur. J. Immunol.* **40**, 699–709 (2010).

23. Higgins, G. T., Wang, J. H., Dockery, P., Cleary, P. E. & Redmond, H. P. Induction of angiogenic cytokine expression in cultured RPE by ingestion of oxidized photoreceptor outer segments. *Invest. Ophthalmol. Vis. Sci.* **44**, 1775–1782 (2003).
24. de Jong, P. T. Age-related macular degeneration. *N. Engl. J. Med.* **355**, 1474–1485 (2006).
25. Binder, C. J. *et al.* Innate and acquired immunity in atherogenesis. *Nature Med.* **8**, 1218–1226 (2002).
26. Flierman, R. & Daha, M. R. The clearance of apoptotic cells by complement. *Immunobiology* **212**, 363–370 (2007).
27. Trouw, L. A. *et al.* C4b-binding protein and factor H compensate for the loss of membrane-bound complement inhibitors to protect apoptotic cells against excessive complement attack. *J. Biol. Chem.* **282**, 28540–28548 (2007).
28. Meri, S. & Pangburn, M. K. Discrimination between activators and nonactivators of the alternative pathway of complement: regulation via a sialic acid/polyanion binding site on factor H. *Proc. Natl Acad. Sci. USA* **87**, 3982–3986 (1990).
29. Leffler, J. *et al.* Annexin-II, DNA, and histones serve as factor H ligands on the surface of apoptotic cells. *J. Biol. Chem.* **285**, 3766–3776 (2010).
30. Savill, J., Dransfield, I., Gregory, C. & Haslett, C. A blast from the past: clearance of apoptotic cells regulates immune responses. *Nature Rev. Immunol.* **2**, 965–975 (2002).
31. Chang, M. K. *et al.* Apoptotic cells with oxidation-specific epitopes are immunogenic and proinflammatory. *J. Exp. Med.* **200**, 1359–1370 (2004).
32. Mhlan, M., Stippa, S., Jozsi, M. & Zipfel, P. F. Monomeric CRP contributes to complement control in fluid phase and on cellular surfaces and increases phagocytosis by recruiting factor H. *Cell Death Differ.* **16**, 1630–1640 (2009).
33. Thakkinian, A. *et al.* Systematic review and meta-analysis of the association between complement factor H Y402H polymorphisms and age-related macular degeneration. *Hum. Mol. Genet.* **15**, 2784–2790 (2006).
34. Donoso, L. A., Vrabec, T. & Kuivaniemi, H. The role of complement factor H in age-related macular degeneration: a review. *Surv. Ophthalmol.* **55**, 227–246 (2010).
35. Clark, S. J. *et al.* His-384 allotypic variant of factor H associated with age-related macular degeneration has different heparin binding properties from the non-disease-associated form. *J. Biol. Chem.* **281**, 24713–24720 (2006).
36. Green, W. R. & Enger, C. Age-related macular degeneration histopathologic studies. The 1992 Lorenz E. Zimmerman Lecture. *Ophthalmology* **100**, 1519–1535 (1993).
37. Sun, M. *et al.* Light-induced oxidation of photoreceptor outer segment phospholipids generates ligands for CD36-mediated phagocytosis by retinal pigment epithelium: a potential mechanism for modulating outer segment phagocytosis under oxidant stress conditions. *J. Biol. Chem.* **281**, 4222–4230 (2006).
38. Kaemmerer, E., Schutt, F., Krohne, T. U., Holz, F. G. & Kopitz, J. Effects of lipid peroxidation-related protein modifications on RPE lysosomal functions and POS phagocytosis. *Invest. Ophthalmol. Vis. Sci.* **48**, 1342–1347 (2007).
39. Goverdhan, S. V. *et al.* Interleukin-8 promoter polymorphism –251A/T is a risk factor for age-related macular degeneration. *Br. J. Ophthalmol.* **92**, 537–540 (2008).
40. Sofat, R. *et al.* Genetic variation in complement factor H and risk of coronary heart disease: eight new studies and a meta-analysis of around 48,000 individuals. *Atherosclerosis* **213**, 184–190 (2010).
41. Scholl, H. P. *et al.* Systemic complement activation in age-related macular degeneration. *PLoS ONE* **3**, e2593 (2008).
42. Binder, C. J. *et al.* Pneumococcal vaccination decreases atherosclerotic lesion formation: molecular mimicry between *Streptococcus pneumoniae* and oxidized LDL. *Nature Med.* **9**, 736–743 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to M. Ozsvar-Kozma for technical assistance, S. Hälbig for performing the surface plasmon resonance analysis, A. Hartmann for purification of CFH variants from patient plasma, C. Mannhalter for help with genotyping, and E. N. Montano. This work was supported by the Austrian Academy of Sciences, a BRIDGE grant from the Austrian Research Promotion Agency, the SFB Lipotox F30 of the Austrian Science Fund (C.J.B.); NIH grants HL088093 (C.J.B., S.T., J.L.W.), RO1 HL086599 (K.H., J.L.W.), EY14005, EY019044 (J.T.H.); the Edward N. & Della L. Thome Memorial Foundation Awards Program in AMD Research, Research to Prevent Blindness (Wilmer Eye institute) (J.T.H.); the Deutsche Forschungsgemeinschaft (P.F.Z., C.S.); the ProRetina Foundation (N.L., P.F.Z., C.S.); the Fondation Leducq (C.J.B., S.T., J.L.W.); the Wynn-Gund Translational Research Acceleration Program Enhanced Research and Clinical Training Award, National Neurovision Research Institute – Foundation Fighting Blindness, Macular Degeneration Research Award, American Health Assistance Foundation (H.P.N.S.); European Commission and the Seventh European Community Framework Program, Marie Curie Intra-European Fellowship (P.C.I.). K.H. was supported by the Scientist Development Grant 0630228N of the American Heart Association. J.T.H. is the Robert Bond Welch Professor. We thank all patients for participation.

Author Contributions D.W. and C.J.B. conceived the project and designed and analysed the experiments with contributions from J.L.W., P.F.Z., J.T.H., G.S.-F. and C.S.; D.W. performed most of the experiments, with contributions from K.H., N.L., K.L.B., M.C., S.T. and H.B.; H.P.N.S. and P.C.I. obtained and provided AMD plasma samples; D.W. and C.J.B. wrote the manuscript with contributions from J.L.W., P.F.Z. and J.T.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.J.B. (christoph.binder@meduniwien.ac.at).

Detectable radio flares following gravitational waves from mergers of binary neutron stars

Ehud Nakar¹ & Tsvi Piran²

Mergers of neutron-star/neutron-star binaries are strong sources of gravitational waves^{1–3}. They can also launch subrelativistic and mildly relativistic outflows^{4–8} and are often assumed to be the sources of short γ -ray bursts⁹. An electromagnetic signature that persisted for weeks to months after the event would strengthen any future claim of a detection of gravitational waves¹⁰. Here we present results of calculations showing that the interaction of mildly relativistic outflows with the surrounding medium produces radio flares with peak emission at 1.4 gigahertz that persist at detectable (submillijansky) levels for weeks, out to a redshift of 0.1. Slower subrelativistic outflows produce flares detectable for years at 150 megahertz, as well as at 1.4 gigahertz, from slightly shorter distances. The radio transient RT 19870422 (ref. 11) has the properties predicted by our model, and its most probable origin is the merger of a compact neutron-star/neutron-star binary. The lack of radio detections usually associated with short γ -ray bursts does not constrain the radio transients that we discuss here (from mildly relativistic and subrelativistic outflows) because short γ -ray burst redshifts are typically >0.1 and the appropriate timescales (longer than weeks) have not been sampled.

Gravitational-wave detectors, and in particular the advanced LIGO and Virgo interferometers, are being constructed now with the goal of detecting gravitational waves from binary neutron-star coalescence at distances up to a few hundred megaparsecs (redshift $z \approx 0.1$)¹². The detection of an accompanying electromagnetic signal would complement these efforts, providing an independent confirmation of the discovery and increasing the detectors' effective sensitivity. The search for such an electromagnetic signal has therefore attracted much interest. The radioactive decay of ejected debris from the merger would drive a short-lived supernova-like event¹³. For example, ejection of 0.01 solar masses ($0.01M_{\odot}$) from a merger at a distance of 300 Mpc would result in a faint optical flare that peaks after ~ 1 day (ref. 14). Finding, and especially identifying, such rare and faint events in the crowded variable optical sky is an extremely challenging task. Other authors have speculated on the production of low-frequency radio signals from the interaction of the neutron stars' magnetic fields^{15–17}. These attempts focused on electromagnetic signals that are contemporaneous with, or follow quickly, the merger and the gravitational waves. Unfortunately, these predictions are highly uncertain.

Here we predict a robust radio signal that peaks several weeks after the merger. Numerical simulations show that compact binary mergers launch energetic subrelativistic and mildly relativistic outflows^{4–8}. Ejection sources include unbound tidal tails, and winds driven by neutrino heating, nucleosynthesis and electromagnetic processes^{18–21}, emerging from the proto-neutron star or from an accretion disk. Overall, almost all merger models find a significant ejection of mass and energy. In binary neutron-star mergers, an ejection of about 10^{50} erg at $(0.1–0.2)c$ (where c is the speed of light) and about 10^{49} erg as faster ejecta is a fairly robust prediction. The outflow from black-hole/neutron-star mergers is less certain, but it is possibly more energetic and faster¹⁸ ($\sim 10^{52}$ erg at $0.5c$).

The interaction of this outflow with the surrounding tenuous matter generates a blast wave. Although the outflow may be highly non-uniform

initially, it becomes spherical rather quickly. We therefore consider a spherical outflow with energy E and an initial velocity $c\beta_i$ that propagates into a medium with a constant density, n . If the outflow is not ultrarelativistic, it propagates at a constant velocity until time t_{dec} when, at a radius R_{dec} , it collects a mass comparable to its own. Time t_{dec} (in days) is given by

$$t_{\text{dec}} = \frac{R_{\text{dec}}}{c\beta_i} \approx 30 E_{49}^{1/3} n_0^{-1/3} \beta_i^{-5/3} \quad (1)$$

Here and in the following, unless stated otherwise, q_x (where q is any parameter) denotes the value of $q/10^x$ in c.g.s. units. At a radius $R > R_{\text{dec}}$, the flow decelerates, assuming a Sedov–Taylor blast wave: $\beta \approx \beta_i (R/R_{\text{dec}})^{-3/2}$.

The blast wave generates magnetic fields and accelerates particles that emit synchrotron radiation. The same microphysics used successfully to model radio emission of type Ibc supernovae^{22,23}, where $\beta \approx 0.2$, and to model late radio emission of γ -ray bursts^{24,25}, where the flow is mildly relativistic, is applicable here. In both cases, the electrons and the magnetic field are found to carry significant fractions of the total internal energy of the shocked gas, $\epsilon_e \approx \epsilon_B \approx 0.1$. The observed spectra reveal a power-law distribution of the electrons' Lorentz factor, γ : $dN/d\gamma \propto \gamma^{-p}$ for $\gamma > \gamma_m = [(p-2)/(p-1)](m_p/m_e)\epsilon_e\beta^2$, where m_p and m_e are the proton and electron masses, respectively, γ_m is the minimal Lorentz factor of the electron's distribution and $p \approx 2–3$.

The radio spectrum is determined by v_m , the synchrotron frequency of electrons with Lorentz factor γ_m , and by v_a , the synchrotron self-absorption frequency (Supplementary Information). The specific flux, F_ν , at a given frequency is strongly suppressed below v_a , and it decreases as $\nu^{-(p-1)/2}$ above v_m and v_a . The signal across the whole spectrum increases at $t < t_{\text{dec}}$. Its behaviour after t_{dec} depends on the relation between the observed frequency, ν_{obs} , and v_m and v_a . The signal peaks at t_{dec} if $\nu_{\text{obs}} > v_{a,\text{dec}}, v_{m,\text{dec}}$ where $v_{a,\text{dec}} \equiv v_a(t_{\text{dec}})$ and $v_{m,\text{dec}} \equiv v_m(t_{\text{dec}})$. Otherwise the signal peaks when $\nu_{\text{obs}} = v_m$ or when $\nu_{\text{obs}} = v_a$, whichever is latest.

The flare characteristics are most sensitive to the initial velocity of the outflow, β_i . Because the brightest radio emission is observed at t_{dec} , a lower value of β_i implies a longer rise time of the radio emission after the merger. Additionally, $v_{m,\text{dec}}, v_{a,\text{dec}}$ and the peak flux at any observed frequency depend strongly on β_i . As mergers are expected to eject an outflow over a range of velocities, we discuss separately below the observed signature of mildly relativistic ($\beta_i \approx 1$) and subrelativistic ($\beta_i \approx 0.2$) ejecta.

A mildly relativistic blast wave with canonical parameters produces a synchrotron spectrum with $v_{a,\text{dec}} \leq v_{m,\text{dec}} \approx 1$ GHz. The strongest signal is then expected at time t_{dec} (a few weeks after the merger) and around 1.4 GHz (Supplementary Information): the peak of the observed specific flux at ν_{obs} in units of millijanskys is

$$F_{\nu_{\text{obs}},\text{peak}}[\nu_{\text{obs}} > v_{m,\text{dec}}, v_{a,\text{dec}}] \approx 0.3 E_{49} n_0^{\frac{p+1}{4}} \epsilon_{B,-1}^{\frac{p+1}{4}} \epsilon_{e,-1}^{p-1} \beta_i^{\frac{5p-7}{2}} d_{27}^{-2} \left(\frac{\nu_{\text{obs}}}{1.4} \right)^{-\frac{p-1}{2}} \quad (2)$$

¹Raymond and Beverly Sackler School of Physics & Astronomy, Tel Aviv University, Tel Aviv 69978, Israel. ²Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel.

where d is the distance to the merger (here ν_{obs} is in GHz). The peak flux at lower frequencies (<1 GHz) is significantly lower and it is observed at a later time. If the outflow is subrelativistic ($\beta_1 \approx 0.1$ – 0.2), then $\nu_{\text{a,dec}}, \nu_{\text{m,dec}} \leq 150$ MHz and equation (2) is applicable also in the frequency range of low-frequency radio detectors. The flux peaks at t_{dec} , which is of the order of years, and it is brighter at 150 MHz than at 1 GHz by about an order of magnitude. Note that over the whole expected range of blast wave parameters, $\nu_{\text{a}} \leq 1$ GHz at all times and the spectrum above 1 GHz is optically thin during the entire evolution. We stress that in radio supernovae, the surrounding dense winds lead at early time to an optically thick spectrum at $\nu_{\text{obs}} > 1$ GHz, and the transition $\nu_{\text{a}} = \nu_{\text{obs}}$ determines the time and flux at the peak. As discussed below, this different spectral signature enables us to distinguish between merger flares and radio supernovae.

The circum-merger density also strongly affects the flare signature. For example, if the surrounding particle density is $\sim 10^{-3} \text{ cm}^{-3}$, the peak flux from a mildly relativistic ejecta decreases to the microjansky level at a distance of 300 Mpc, and the timescale increases by a factor of ten, to a year. A merger taking place in such a density can be detected only up to distances of ~ 100 Mpc (Table 1.) The density is expected to vary significantly, from $n \approx 1 \text{ cm}^{-3}$, in galactic disks, to $n \approx 10^{-6} \text{ cm}^{-3}$, for mergers taking place outside their host galaxies. Because all observed Galactic neutron-star binaries reside within the Galactic disk, where the average density is $n \approx 1 \text{ cm}^{-3}$, a significant fraction of the cosmological mergers are expected also to take place in rather dense environments. We therefore use $n = 1 \text{ cm}^{-3}$ as the canonical density value. If mergers produce short γ -ray bursts, then observations of their afterglows support this value (Supplementary Information).

An intriguing possibility is that compact merger events also eject ultrarelativistic jets that produce short γ -ray bursts⁹ (SGRBs). It is important to examine the relationship between SGRBs and the radio flares discussed here, assuming that mergers are producing SGRBs. An SGRB beamed towards us will be observed in coincidence with the gravitational-wave signal, providing a clear electromagnetic counterpart. Even if the SGRB itself is missed, owing to partial sky coverage, its afterglow will be easily detectable. However, SGRBs are expected to be beamed, and only rarely will one point towards us. A beamed SGRB observed off-axis produces, once it has slowed down, a long-lasting radio ‘orphan’ afterglow²⁶, similar in its characteristics to the mildly relativistic signal discussed above. However, the total energy (corrected for beaming) in the ultrarelativistic jet is at most comparable to—and probably lower than—that of the mildly relativistic ejecta. Consequently the latter will dominate the radio emission.

The radio remnant signals that we consider here, which are generated by subrelativistic and mildly relativistic outflows, could not have been detected in the radio afterglow searches that were carried out following SGRB triggers. The reasons are twofold. First, SGRBs are typically detected at distances of 1–3 Gpc, far beyond the detection horizon of gravitational-wave detectors. Hence the signals are much weaker than those associated with detected gravitational-wave events. Second, SGRB afterglow searches are optimized to detect the emission from ultrarelativistic ejecta pointing towards the observer. Such emission

peaks at higher frequencies and on shorter timescales than the emission from mildly and subrelativistic ejecta discussed here. SGRB afterglow searches are done at a sensitivity of ~ 0.1 mJy at 4.8–8.5 GHz during the first week or two after the bursts (see, for example, refs 27–29). Equation (2) implies that over the distance range 1–3 Gpc, the radio signal of mildly relativistic ejecta with energy 10^{50} erg that propagate into a medium of density $n = 1 \text{ cm}^{-3}$ peaks after ~ 60 days at a flux of ~ 0.01 – 0.1 mJy at 5 GHz. The flux before the peak rises as $(t/t_{\text{dec}})^3$ (Supplementary Information). Therefore, these early radio afterglow searches could not have detected the radio signal, even if the mildly relativistic ejecta had an energy of 10^{51} erg. Thus, the paucity of detected SGRB radio afterglows has no direct implication for the nature of the remnants we discuss here.

A new wave of radio detectors is now coming online. The most sensitive operate at frequencies of 1.4 GHz and higher. Table 1 summarizes the relevant properties of these facilities and their detection horizons. The best facility for a targeted search, following a detection of a candidate gravitational-wave source, is clearly the EVLA. A deep, ~ 50 μJy , localized EVLA search of the 10–100 deg² error box of a gravitational-wave trigger³⁰ can detect mildly relativistic ejecta (with an energy of even $\sim 10^{48}$ erg) out to the horizon of advanced gravitational-wave detectors. The upcoming lower-frequency LOFAR sensor array will be more effective in searches for subrelativistic outflows, whose signals peak at LOFAR’s frequencies, thus compensating for LOFAR’s lower sensitivity. LOFAR is also relatively more effective when searching for flares in a low-density medium.

Even before the completion of the advanced gravitational-wave detectors, blind searches can identify radio flares from compact binary mergers. Identification of radio emission from any merger type (binary neutron star or black hole/neutron star) would determine the merger rate, which is a parameter of utmost importance for the design and operation of advanced detectors. With a merger rate of $300 \text{ Gpc}^{-3} \text{ yr}^{-1}$, we expect these facilities to detect ~ 20 remnants from mildly relativistic outflows with an energy of $E = 10^{49}$ erg (and $\sim 1,000$ remnants if $E = 10^{50}$ erg), in a single whole-sky snapshot. LOFAR may detect a dozen transients in a whole-sky survey, even if only a subrelativistic outflow with an energy of 10^{50} erg is ejected (see Supplementary Information for details, as well as for a discussion of ways to distinguish these flares from other possible radio transients).

Remarkably, the observed 5 GHz transient RT 19870422 (ref. 11) shows all the expected properties of the radio remnant of a compact binary merger. At 1 Gpc distance and with a duration of two months, this transient is what we would expect from a mildly relativistic outflow with an energy of $\sim 10^{50}$ erg. The inferred rate of similar transients¹¹, 80 – $20,000 \text{ Gpc}^{-3} \text{ yr}^{-1}$, is fully consistent with the estimates of compact binary mergers. This transient is therefore an excellent candidate to be the first observed radio remnant of a merger. Unfortunately, we cannot rule out the possibility that this is an especially bright radio supernova¹¹. We note, however, that this latter interpretation requires a supernova brighter by an order of magnitude than any radio supernovae previously observed. Simultaneous optical observations or multiwavelength radio observations could have easily distinguished

Table 1 | Observing radio flares

Radio facility	Observing frequency (GHz)	Field of view (deg ²)	One-hour r.m.s.* (μJy)	One-hour detection horizon†			Ten-hour detection horizon‡
				$\beta_1 \approx 1$, $E_{49} = 1$, $n_0 = 1$	$\beta_1 \approx 1$, $E_{49} = 10$, $n_0 = 1$	$\beta_1 = 0.2$, $E_{49} = 10$, $n_0 = 1$, $p = 2.5$	$\beta_1 \approx 1$, $E_{49} = 1$, $n_0 = 10^{-3}$, $p = 2$
EVLA	1.4	0.25	7	1 Gpc	3.3 Gpc	370 Mpc	140 Mpc
ASKAP	1.4	30	30	500 Mpc	1.6 Gpc	180 Mpc	70 Mpc
MeerKAT	1.4	1.5	35	500 Mpc	1.6 Gpc	165 Mpc	65 Mpc
Apertif	1.4	8	50	400 Mpc	1.25 Gpc	140 Mpc	50 Mpc
LOFAR	0.15	20	1,000	35 Mpc	90 Mpc	70 Mpc	20 Mpc

Shown are properties and detection horizons (neglecting cosmological corrections) for an observation at different radio facilities of blast waves with various values of β_1 , E_{49} , n_0 and p (in all cases, $\epsilon_{\text{e}} \approx \epsilon_{\text{B}} \approx 0.1$; see text for definitions of these symbols). Information on facilities is available as follows: EVLA (<http://www.aoc.nrao.edu/evla/>); ASKAP (<http://www.atnf.csiro.au/projects/askap/technology.html>); MeerKAT (<http://www.ska.ac.za/meerkat/>); Apertif (<http://www.astron.nl/general/apertif/apertif>); and LOFAR (<http://lofar.org>).

* The root mean squared value of the background noise for one hour of observation.

† The distance at which the observed peak flux is four times the one-hour r.m.s.

‡ The distance at which the observed peak flux is four times the root mean squared value of the background noise for ten hours of observation.

between the two possibilities. Unfortunately, no such observations are available. However, the detection rate implied by this event is very high, indicating that similar events could easily be detected by a relatively small-scale survey and that their nature should be easily probed.

Received 11 February; accepted 11 July 2011.

Published online 28 September 2011.

1. Taylor, J. H. & Weisberg, J. M. A new test of general relativity — gravitational radiation and the binary pulsar PSR 1913+16. *Astrophys. J.* **253**, 908–920 (1982).
2. Esposito, L. W. & Harrison, E. R. Properties of the Hulse-Taylor binary pulsar system. *Astrophys. J.* **196**, L1–L2 (1975).
3. Wagoner, R. V. Test for the existence of gravitational radiation. *Astrophys. J.* **196**, L63–L65 (1975).
4. Rosswog, S., Davies, M. B., Thielemann, F. & Piran, T. Merging neutron stars: asymmetric systems. *Astron. Astrophys.* **360**, 171–184 (2000).
5. Ruffert, M. & Janka, H. Coalescing neutron stars — a step towards physical models. III. Improved numerics and different neutron star masses and spins. *Astron. Astrophys.* **380**, 544–577 (2001).
6. Yamamoto, T., Shibata, M. & Taniguchi, K. Simulating coalescing compact binaries by a new code (SACRA). *Phys. Rev. D* **78**, 064054 (2008).
7. Rezzolla, L., Baiotti, L., Giacomazzo, B., Link, D. & Font, J. A. Accurate evolutions of unequal-mass neutron-star binaries: properties of the torus and short GRB engines. *Class. Quantum Grav.* **27**, 114105 (2010).
8. Kiuchi, K., Sekiguchi, Y., Shibata, M. & Taniguchi, K. Exploring binary-neutron-star-merger scenario of short-gamma-ray bursts by gravitational-wave observation. *Phys. Rev. Lett.* **104**, 141101 (2010).
9. Eichler, D., Livio, M., Piran, T. & Schramm, D. N. Nucleosynthesis, neutrino bursts and gamma-rays from coalescing neutron stars. *Nature* **340**, 126–128 (1989).
10. Kochanek, C. S. & Piran, T. Gravitational waves and gamma-ray bursts. *Astrophys. J.* **417**, L17–L20 (1993).
11. Bower, G. C. *et al.* Submillijansky transients in archival radio observations. *Astrophys. J.* **666**, 346–360 (2007).
12. Smith, J. R. & LIGO Scientific Collaboration. The path to the enhanced and advanced LIGO gravitational-wave detectors. *Class. Quantum Grav.* **26**, 114013 (2009).
13. Li, L. & Paczyński, B. Transient events from neutron star mergers. *Astrophys. J.* **507**, L59–L62 (1998).
14. Metzger, B. D. *et al.* Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei. *Mon. Not. R. Astron. Soc.* **406**, 2650–2662 (2010).
15. Hansen, B. M. S. & Lyutikov, M. Radio and X-ray signatures of merging neutron stars. *Mon. Not. R. Astron. Soc.* **322**, 695–701 (2001).
16. Moortgat, J. & Kuijpers, J. Gravitational waves in magnetized relativistic plasmas. *Phys. Rev. D* **70**, 023001 (2004).
17. Pshirkov, M. S. & Postnov, K. A. Radio precursors to neutron star binary mergings. *Astrophys. Space Sci.* **330**, 13–18 (2010).
18. Rosswog, S. Mergers of neutron star-black hole binaries with small mass ratios: nucleosynthesis, gamma-ray bursts, and electromagnetic transients. *Astrophys. J.* **634**, 1202–1213 (2005).
19. Levinson, A. General relativistic, neutrino-assisted magnetohydrodynamic winds-theory and application to gamma-ray bursts. I. Schwarzschild geometry. *Astrophys. J.* **648**, 510–522 (2006).
20. Metzger, B. D., Piro, A. L. & Quataert, E. Time-dependent models of accretion discs formed from compact object mergers. *Mon. Not. R. Astron. Soc.* **390**, 781–797 (2008).
21. Dessart, L., Ott, C. D., Burrows, A., Rosswog, S. & Livne, E. Neutrino signatures and the neutrino-driven wind in binary neutron star mergers. *Astrophys. J.* **690**, 1681–1705 (2009).
22. Chevalier, R. A. Synchrotron self-absorption in radio supernovae. *Astrophys. J.* **499**, 810–819 (1998).
23. Chevalier, R. A. & Fransson, C. Circumstellar emission from Type Ib and Ic supernovae. *Astrophys. J.* **651**, 381–391 (2006).
24. Frail, D. A., Waxman, E. & Kulkarni, S. R. A 450 day light curve of the radio afterglow of GRB 970508: fireball calorimetry. *Astrophys. J.* **537**, 191–204 (2000).
25. Frail, D. A. *et al.* Accurate calorimetry of GRB 030329. *Astrophys. J.* **619**, 994–998 (2005).
26. Rhoads, J. E. How to tell a jet from a balloon: a proposed test for beaming in gamma-ray bursts. *Astrophys. J.* **487**, L1–L4 (1997).
27. Soderberg, A. M. & Frail, D. A. GRB060313: radio observations. *GCN Circ.* **4884** (2006).
28. van der Horst, A. J. GRB 060801: WSRT radio observations. *GCN Circ.* **5390** (2006).
29. Wieringa, M. H., Chandra, P. & Frail, D. A. ATCA radio observations of short hard burst GRB 071112B. *GCN Circ.* **7095** (2007).
30. Wen, L. & Chen, Y. Geometrical expression for the angular resolution of a network of gravitational-wave detectors. *Phys. Rev. D* **81**, 082001 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Frail, S. Kulkarni, A. Levinson, A. MacFadyen, E. Ofek and S. Rosswog for discussions. This research was supported by an ERC advanced research grant, by the Israeli Center for Excellence for High Energy Astrophysics, by the ISF (grant No. 174/08) and by an IRG grant.

Author Contributions Both authors contributed equally.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.P. (tsvi@phys.huji.ac.il) or E.N. (udini@wise.tau.ac.il).

Resonances arising from hydrodynamic memory in Brownian motion

Thomas Franosch^{1*}, Matthias Grimm^{2,3*}, Maxim Belushkin⁴, Flavio M. Mor³, Giuseppe Foffi⁴, László Forró³ & Sylvia Jeney^{2,3}

Observation of the Brownian motion of a small probe interacting with its environment provides one of the main strategies for characterizing soft matter^{1–4}. Essentially, two counteracting forces govern the motion of the Brownian particle. First, the particle is driven by rapid collisions with the surrounding solvent molecules, referred to as thermal noise. Second, the friction between the particle and the viscous solvent damps its motion. Conventionally, the thermal force is assumed to be random and characterized by a Gaussian white noise spectrum. The friction is assumed to be given by the Stokes drag, suggesting that motion is overdamped at long times in particle tracking experiments, when inertia becomes negligible. However, as the particle receives momentum from the fluctuating fluid molecules, it also displaces the fluid in its immediate vicinity. The entrained fluid acts back on the particle and gives rise to long-range correlations^{5,6}. This hydrodynamic ‘memory’ translates to thermal forces, which have a coloured, that is, non-white, noise spectrum. One hundred years after Perrin’s pioneering experiments on Brownian motion^{7–9}, direct experimental observation of this colour is still elusive¹⁰. Here we measure the spectrum of thermal noise by confining the Brownian fluctuations of a microsphere in a strong optical trap. We show that hydrodynamic correlations result in a resonant peak in the power spectral density of the sphere’s positional fluctuations, in strong contrast to overdamped systems. Furthermore, we demonstrate different strategies to achieve peak amplification. By analogy with microcantilever-based sensors^{11,12}, our results reveal that the particle–fluid–trap system can be considered a nanomechanical resonator in which the intrinsic hydrodynamic backflow enhances resonance. Therefore, instead of being treated as a disturbance, details in thermal noise could be exploited for the development of new types of sensor and particle-based assay in lab-on-a-chip applications^{13,14}.

Einstein’s theory of Brownian motion¹⁵, published in 1905, received considerable attention and was later reformulated in terms of a Langevin equation¹⁶. In it, particle motion is driven by thermal fluctuations induced through collisions with the fluid molecules. These rapid ‘kicks’ are assumed to be random and independent at frequencies much smaller than the collision rate of ~ 1 THz. The thermal force consequently has a white noise spectrum¹⁶; that is, the spectrum is constant over a wide range of frequencies. Momentum is transferred from the particle to the fluid at times $\tau_p = m_p/\gamma$ (Fig. 1a, left), where $\gamma = 6\pi\eta R$ is the coefficient of static friction of the particle for macroscopic no-slip boundary conditions, m_p is the particle’s mass, η is the shear viscosity of the fluid and R is the radius of the particle (which is taken to be spherical). However, when the densities of the particle and the fluid, ρ_p and, respectively, ρ_f , are comparable, their coupling becomes important^{17,18}. As the suspended particle fluctuates through the solvent, long-range correlations build up as a result of momentum exchange, leading to hydrodynamic memory in the solvent. Hence, an additional timescale, $\tau_f = R^2\rho_f/\eta$, which describes the time needed by the perturbed fluid flow field to

diffuse over one particle radius (Fig. 1a, middle), becomes important. According to the fluctuation–dissipation theorem, the statistics of the thermal force, $F_{th}(t)$, is characterized by a delta-correlated white noise term and a coloured, frequency-dependent component that reflects the retarded viscous response of the fluid continuum to the particle.

To measure directly the predicted correlations in thermal noise, we combined strong optical trapping with high-resolution, 3D position detection¹⁹ (Supplementary Information, section 1). The resulting force balance for the particle reads $m_p\ddot{x}(t) = F_{fr}(t) - Kx(t) + F_{th}(t)$, where $x(t)$ is the particle’s displacement from the trap centre (with a dot denoting differentiation with respect to time), $F_{fr}(t)$ is the non-instantaneous friction force on the particle and K is the stiffness of the optical trap. This harmonic restoring force gives the trap relaxation time of $\tau_K = \gamma/K$ (Fig. 1a, right). At long times, strong trapping eventually dominates over friction and becomes the main force counteracting thermal excitation. The Langevin equation reduces then to $Kx(t) \approx F_{th}(t)$. Consequently, when tracking the fluctuating motion of the particle in a strong harmonic potential (Fig. 1b, c), we effectively probe the thermal force of the fluid¹⁰. Correlations in thermal noise

$$\langle F_{th}(t)F_{th}(0) \rangle \approx K^2 \langle x(t)x(0) \rangle \quad (1)$$

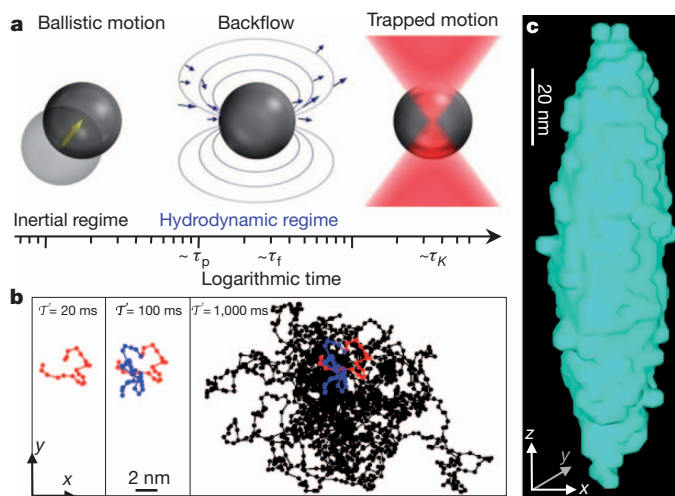


Figure 1 | Characteristic time scales of a Brownian particle confined by the three-dimensional (3D) harmonic potential of an optical trap. **a**, On very short timescales ($t \ll \tau_p$), the particle undergoes ballistic motion governed by m_p (left). On timescale τ_p hydrodynamic backflow develops (centre; solid lines show the emerging fluid velocity field; arrows are obtained from our computer simulations). Finally, for $t \gtrsim \tau_K$, the harmonic potential of the trap sets in and confines particle diffusion (right). **b**, Trajectories of the trapped sphere measured at three different time intervals, in the dimensions, x and y , lateral to the optical axis, z . **c**, 3D position histogram of the same sphere after the measurement time $T \approx 4$ ms. The small displacements of the bead are indicative of the strong trapping forces.

¹Institut für Theoretische Physik, Friedrich-Alexander-Universität Erlangen-Nürnberg, Staudtstraße 7, 91058 Erlangen, Germany. ²M. E. Müller Institute for Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, Basel 4056, Switzerland. ³Laboratory of Physics of Complex Matter, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. ⁴Institute of Theoretical Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

*These authors contributed equally to this work.

become directly accessible through the positional autocorrelation function, $\text{PAF}(t) = \langle x(t)x(0) \rangle$, calculated from the recorded fluctuations of the trapped sphere (Fig. 1b).

The sequential order and magnitudes of the various timescales depend on the size and mass of the Brownian particle, the nature of the solvent and the stiffness of the optical trap. The experimental challenge consists of exploring a wide range of timescales. Therefore, we optimized our set-up²⁰ to achieve a resolution of ~ 1 nm in space and close to $1 \mu\text{s}$ in time (Supplementary Information, section 1). In typical optical tweezers experiments, silica or polystyrene spheres immersed in water and with sizes $\lesssim 1 \mu\text{m}$ are used. This yields values for τ_f and τ_p of less than $1 \mu\text{s}$, which is below our temporal resolution limit²¹. Therefore, we instead used melamine resin beads with diameters of between 2 and $3 \mu\text{m}$ and suspended them in acetone, which is three times less viscous than water. In this set-up, $\tau_p \approx 1\text{--}3 \mu\text{s}$ and $\tau_f \approx 2\text{--}6 \mu\text{s}$. Furthermore, the difference between the refractive indices of resin ($n = 1.68$) and acetone ($n = 1.36$) was high enough to provide good trapping efficiency. With such an experimental configuration, we could increase τ_f and K sufficiently to bring the trap relaxation time, τ_K , close to τ_f . This made the window in which mainly thermal force correlations determine the bead's dynamics (Fig. 1a, middle) experimentally accessible. On these timescales, the mass of the particle is already negligible, leading to a clear separation between the inertial and hydrodynamic regimes of Brownian motion (Supplementary Information, section 5).

Figure 2a, b shows the mean squared displacements and positional autocorrelation functions calculated from the measured position fluctuations, $x(t)$ (Supplementary Information, section 3), of a single resin sphere immersed in water (green circles) or in acetone (blue circles) and held with comparable optical forces. $\text{PAF}(t)$ has a clear zero-crossing followed by anticorrelations. The appearance of anticorrelations is in remarkable contrast with the exponential relaxation, $(k_B T/K)e^{-t/\tau_K}$, characteristic of overdamped harmonic oscillators subject to instantaneous Stokes friction $-\gamma\dot{x}(t)$. In the frequency domain (Fig. 2c), the corresponding power spectral density, $\text{PSD}(f)$, shows that increasing τ_f and hence decreasing the ratio τ_K/τ_f by reducing the fluid's viscosity, caused the emergence of a resonance. This resonant peak indicates that the thermal force spectrum is enhanced as frequency increases (blue circles). In water (green circles), the maximal corner frequency, $f_K = 1/2\pi\tau_K$, we obtained resulted in an enhancement of the PSD close to our noise limit. Nevertheless, deviations from the simple Lorentzian, $\text{PSD}(f) = 2(k_B T/K)\tau_K/[1 + (2\pi f\tau_K)^2]$, of overdamped systems (green line) are clearly visible for frequencies around f_K .

For a quantitative description, we solve a Langevin equation with no-slip boundary conditions accounting for slow vortex diffusion^{17,18} and trapping^{22,23}. Our data (Fig. 2, symbols) are in excellent agreement with the theoretical expression (Fig. 2, black lines) over three decades in time and four orders of magnitude in signal (Fig. 2b), and we observe a hydrodynamic power-law tail, $\text{PAF}(t) \approx -(k_B T\gamma/K^2)\sqrt{\tau_f/4\pi t^3}$, which by equation (1) directly reflects the corresponding persistent correlations in the thermal forces

$$\langle F_{\text{th}}(t)F_{\text{th}}(0) \rangle = -k_B T\gamma\sqrt{\tau_f/4\pi t^{-3/2}}$$

for times for which compressibility effects from the fluid can be ignored (Supplementary Information, section 4). The negative overshoot in the PAF and the equivalent resonant peak in the PSD originate solely from the hydrodynamic coupling between the fluid and the particle.

In the Fourier domain, positional fluctuations $\hat{x}(f) = \int x(t)e^{2\pi ift} dt$ are connected to the thermal forces by a linear relation, $\hat{x}(f) = \hat{G}(f)\hat{F}_{\text{th}}(f)$, where $\hat{G}(f)$ is the Fourier transform of the Green function (Supplementary Information, section 4). Consequently, $\hat{x}(f)$ is a filtered signal of the noise $\hat{F}_{\text{th}}(f)$, and the PSD of the Brownian thermal noise, $\text{PSD}_{\text{th}}(f)$, is related to the measured PSD by

$$\text{PSD}_{\text{th}}(f) = |\hat{G}(f)|^{-2} \text{PSD}(f) \quad (2)$$

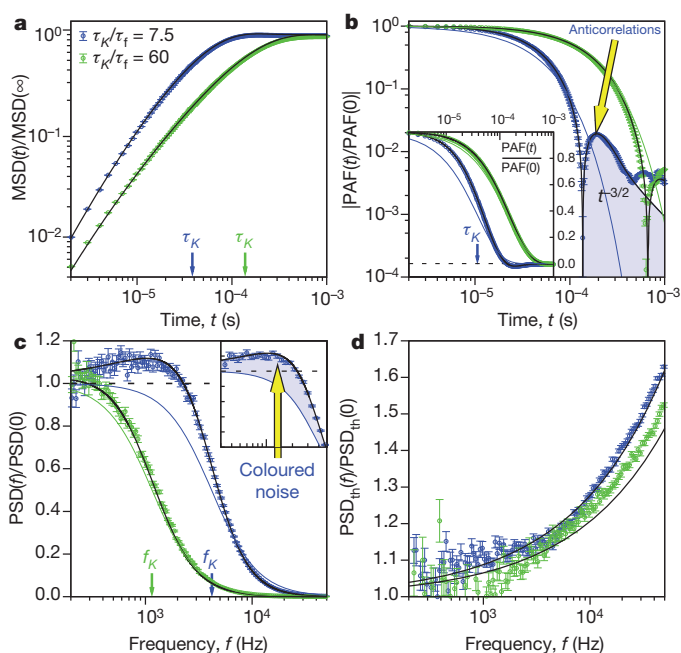


Figure 2 | The colour of thermal force. **a**, Double-logarithmic plot of the mean squared displacement (MSD) normalized to its long-time limit, $2k_B T/K$ (k_B , Boltzmann's constant), for an optically trapped ($K \approx 205 \mu\text{N m}^{-1}$), melamine resin sphere ($R = 1.45 \mu\text{m}$) in water (green circles: $\tau_f = 2.3 \mu\text{s}$, $\tau_K = 138.0 \mu\text{s}$) or acetone (blue circles: $\tau_f = 5.1 \mu\text{s}$, $\tau_K = 38.3 \mu\text{s}$). **b**, Double-logarithmic plot of the corresponding PAF(t) blocked in 100 bins per decade and normalized to its initial value, $\text{PAF}(0) = k_B T/K$. The persistent anticorrelations are visible after the zero-crossing (narrow spike) and follow a $t^{-3/2}$ power-law decay. The green and blue lines indicate exponential relaxations and serve as guide to the eye. Inset, log-linear plot of the same data. **c**, Log-linear representation of the corresponding PSD blocked in 50 bins per decade and normalized to its zero-frequency value, $2k_B T\tau_K/K$. The green and blue lines are Lorentzian spectra for reference. Inset, magnified view of the enhancement of the PSD, blocked in 20 bins per decade, reflecting the colour of thermal noise. **d**, Direct representation of the PSD of F_{th} (equation (2)). The black lines correspond to the full hydrodynamic theory including inertial effects^{22,23}. The parameters τ_f and τ_K were extracted from the fit to the theory. Error bars, 1 s.e. of the mean from blocking.

The data shown in Fig. 2d confirm the departure from white noise through a drastic increase in thermal noise at higher frequencies. Deviations from Gaussian white noise are towards the blue end of the spectrum at frequencies that are much smaller than the collision rate of the solvent molecules, and reflect the colour of thermal force¹⁰.

The observed resonance in the PSD can be enhanced by decreasing the ratio τ_K/τ_f and hence increasing K . Figure 3a shows peak amplification with increasing laser power up to a stiffness of $412 \mu\text{N m}^{-1}$. In the hydrodynamic regime, Brownian motion is strongly sensitive to particle size because the determinant timescale, τ_f , is proportional to R^2 . A difference in the bead radius, ΔR , of only a few per cent results in a detectable shift in the PAF around its zero-crossing, in acetone as well as in water (Fig. 3b).

Experimental access to short timescales reveals a resonance in Brownian motion where overdamped motion is commonly assumed. For a given solvent and particle, it is possible to investigate the dynamics of the system in different regimes by decreasing τ_K and detecting the position fluctuations at the highest bandwidth. Stronger and narrower resonances can be obtained in the inertial regime, where Brownian motion is also sensitive to the particle's mass^{24,25}. To reach this window, τ_K has to be brought close to τ_p by increasing K or m_p . Although heavier particles, which simultaneously allow for more efficient trapping, are still to be developed²⁶, the timeline displayed in Fig. 1a can be explored theoretically and by means of computer simulations. Transition to the inertial regime is marked by the appearance of a peak in the case of the

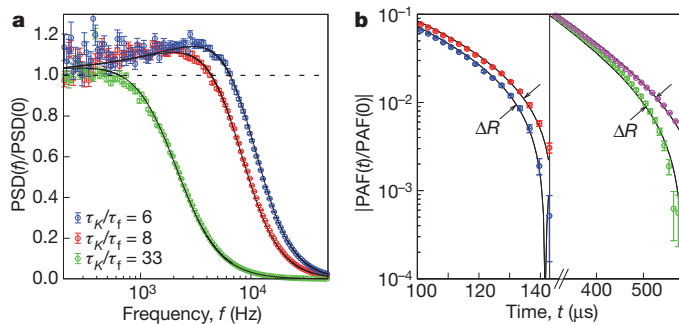


Figure 3 | Enhancing resonance and sensitivity to particle size. **a**, Log-linear representation of the normalized PSD, blocked in 50 bins per decade, of a resin sphere ($R = 1 \mu\text{m}$) in acetone for increasing trap stiffness (green, $K = 77 \mu\text{N m}^{-1}$; red, $309 \mu\text{N m}^{-1}$; blue, $412 \mu\text{N m}^{-1}$). **b**, Magnified view of a log-linear plot of the normalized PAF(t), blocked in 100 bins per decade, close to the zero-crossing for resin spheres of slightly different radii held by traps of the same stiffness, in acetone (blue, $R = 1.45 \mu\text{m}$; red, $1.35 \mu\text{m}$; $K \approx 205 \mu\text{N m}^{-1}$) and water (green, $1.50 \mu\text{m}$; magenta, $1.45 \mu\text{m}$; $K \approx 195 \mu\text{N m}^{-1}$). The black lines in each plot correspond to the full hydrodynamic theory^{22,23}. Data were acquired and processed as described in Fig. 2. Error bars, 1 s.e. of the mean from blocking.

harmonic oscillator with $\tau_K < 2\tau_p$. When τ_K is further decreased, the mass term in the Langevin equation eventually becomes larger than the friction term. Interestingly, in comparison with the simple harmonic oscillator (Fig. 4a, dashed lines), the peak is significantly enhanced and its position, f_{max} , is shifted to lower frequencies by the contribution of hydrodynamic memory (Fig. 4a, solid lines).

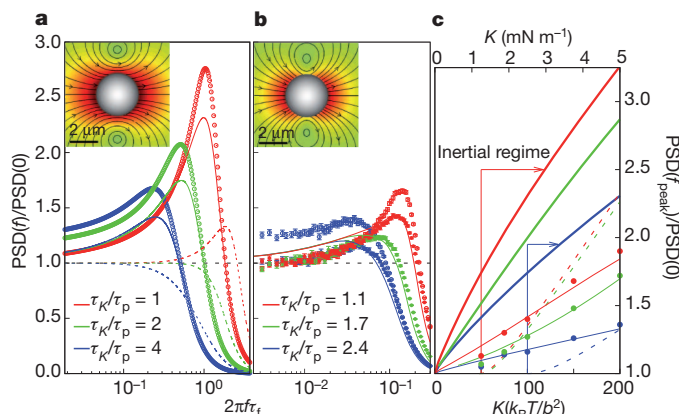


Figure 4 | Transition to the inertial regime. **a**, Theoretical PSD of a resin sphere with no-slip boundary conditions in acetone ($R = 1.5 \mu\text{m}$, $\tau_p = 2.3 \mu\text{s}$, $\tau_f = 5.4 \mu\text{s}$), for very strong traps (blue, $K = 1.02 \text{ mN m}^{-1}$; green, 2.04 mN m^{-1} ; red, 4.07 mN m^{-1}). The dashed lines show the results for a damped harmonic oscillator¹⁶, for comparison. The coloured circles represent the corresponding $\text{PSD}_{\text{exc}}(f)$ with $g = 50\%$ and $f_{\text{exc}} \approx 2f_{\text{peak}}$. Inset, corresponding fluid velocity field developing around the sphere when moving in the x direction at $f = 1/2\pi\tau_f$. The arrows indicate the direction of the velocity field. **b**, Simulation data (filled circles) for an equivalent system with $\tau_p = 0.7\tau_f$ evaluated in a compressible fluid under full-slip conditions, $\gamma = 4\pi\eta R$. The coloured theoretical lines account for vortex diffusion, as well as for sound waves. The open circles are simulation data for an excitation with $g = 50\%$ and $f_{\text{exc}} = 2f_{\text{max}}$. Inset, corresponding fluid velocity field at $f = 1/2\pi\tau_f$, decreasing from red to green. All curves in **a** and **b** are normalized to the zero-frequency value of the respective non-excited PSD. **c**, Normalized peak height for increasing trap strength, calculated for particles of different sizes and densities in acetone, where b is the unit length of simulations (Supplementary Information, Section 7) (thick lines; blue: $2b = R = 1.0 \mu\text{m}$, $\rho_p = 1,510 \text{ kg m}^{-3}$; green: $R = 1.0 \mu\text{m}$, $\rho_p = 3,020 \text{ kg m}^{-3}$; red: $R = 2.0 \mu\text{m}$, $\rho_p = 1,510 \text{ kg m}^{-3}$) and compared with the harmonic oscillator (dashed lines). Simulation data (symbols) are compared to the corresponding theory (thin lines).

The study of Brownian motion on short timescales in a medium with hydrodynamic effects has also become accessible by advanced simulation techniques. We used multiparticle collision dynamics (see ref. 27 and references therein) with molecular dynamics coupling between solute and solvent particles. The method yielded a compressible solvent and correctly reflected the hydrodynamic effects at coarse-grained scales. The approach was implemented most conveniently for full-slip boundary conditions at the solute–solvent interface, whereas our experiments obeyed no-slip conditions. The simulation results for the PSD show that a resonance emerges in the hydrodynamic regime as well as in the inertial regime (Fig. 4b, filled circles), irrespective of boundary conditions at the solvent–solute interface. However, the weak coupling between the Brownian particle and the surrounding fluid yielded a resonance much weaker than that which emerged under no-slip conditions (Fig. 4a). The collected data follow the theoretical curves (Fig. 4b, coloured lines), where friction is evaluated for a compressible fluid under full-slip conditions (Supplementary Information, sections 4 and 7).

The enhanced resonance is sensitive to the size of the bead in the hydrodynamic regime (Fig. 4c, red curves versus green and blue curves). Also, it is mass sensitive in the inertial regime of the simulation and, more markedly, under experimental conditions. In contrast, for the harmonic oscillator, sensitivity to particle size is much lower and sensitivity to its mass occurs only in the underdamped regime, where $\tau_K < 2\tau_p$ (Fig. 4c, dashed lines).

Additional peak amplification can be achieved through parametric resonance^{25,28,29} by periodically modulating the trap strength at frequency f_{exc} : $K(t) = K[1 + g\cos(2\pi f_{\text{exc}}t)]$. We obtained the theoretical excited PSD, $\text{PSD}_{\text{exc}}(f)$, normalized to the initial value of the non-excited PSD, from the solution of a parametrically modulated Langevin equation, including hydrodynamic memory, using second-order perturbation theory in the reduced modulation amplitude, g (Supplementary Information, section 6). As for a harmonic oscillator, also in the presence of coloured friction, the greatest additional peak amplification is achieved at a frequency of $f_{\text{exc}} \approx 2f_{\text{peak}}$, which yields an increase of up to 20% when $g = 50\%$ (Fig. 4a, open circles). Comparable results were obtained with computer simulations (Fig. 4b, open circles).

Exploiting the hydrodynamic and inertial regimes of Brownian motion for particle-based assays will become a common technological approach^{13,14}. We anticipate that changes in the particle's morphology, such as swelling or a reaction occurring at its surface, will alter short-time dynamics and become detectable. As single cells, microorganisms and microcarriers can also be bound harmonically^{12,30}, short-time detection of their Brownian fluctuations may become a sensitive way to characterize their state or evolution in native solutions and without specific markers. Reciprocally, changes in the medium surrounding the probing particle modulate the particle's fluctuation spectrum^{2–4}, offering a means of studying dynamic polymer systems in great detail.

METHODS SUMMARY

Melamine resin microspheres ($\rho_p = 1,510 \text{ kg m}^{-3}$, $R = 1.5, 1.45, 1.35$ or $1 \mu\text{m}$) were suspended in high-purity acetone ($\eta = 0.32 \text{ cP}$ ($1 \text{ cP} = 1 \text{ mPa s}$), $\rho_f = 790 \text{ kg m}^{-3}$) or water ($\eta = 0.95 \text{ cP}$, $\rho_f = 1,000 \text{ kg m}^{-3}$) at minimal concentrations to allow trapping and observation of a single particle. After loading, the sample chamber was mounted onto the 3D piezo stage of our custom-made inverted microscope/optical trap set-up²⁰. A bead was trapped in the focus of a Gaussian trapping beam produced by a diode-pumped, ultralow-noise Nd:YAG laser with a wavelength of $\lambda = 1,064 \text{ nm}$ and a maximal output power of 500 mW in continuous-wave mode. At the focus, the remaining power was measured to be 200 mW for the stiffest traps used. To avoid surface effects, the trapped bead was brought by the piezo stage no closer than $40 \mu\text{m}$ to the top or bottom glass surface of our sample chambers, which were more than $100 \mu\text{m}$ thick. Fluctuations in the position of the bead were detected in 3D by an InGaAs quadrant photodiode³¹ with a diameter of 2.0 mm. The signals from the quadrant photodiode were fed into a custom-built preamplifier, which provided two differential signals between the photodiode quadrants, giving the fluctuations in the x and y directions, and one signal that is proportional to the

total light intensity, yielding the fluctuation in the direction parallel to the optical axis, z . Subsequently, we used differential amplifiers to adjust the preamplifier signals for optimal digitalization by the data acquisition board with a dynamic range of 12 bits. All data were collected for $T \approx 50$ s at a sampling rate of 1 MHz, corresponding to $\sim 5 \times 10^7$ data points. The PSD presented in Figs 2 and 3 were computed from overlapping windows of 2^{22} points.

Received 31 January; accepted 18 August 2011.

- Mason, T. G. & Weitz, D. A. Optical measurements of frequency-dependent linear viscoelastic moduli of complex fluids. *Phys. Rev. Lett.* **74**, 1250–1253 (1995).
- Nowak, A. P. *et al.* Rapidly recovering hydrogel scaffolds from self-assembling diblock copolypeptide amphiphiles. *Nature* **417**, 424–428 (2002).
- Gardel, M. L. *et al.* Elastic behavior of cross-linked and bundled actin networks. *Science* **304**, 1301–1305 (2004).
- Chaudhuri, O., Parekh, S. H. & Fletcher, D. A. Reversible stress softening of actin networks. *Nature* **445**, 295–298 (2007).
- Alder, B. J. & Wainwright, T. E. Velocity autocorrelations for hard spheres. *Phys. Rev. Lett.* **18**, 988–990 (1967).
- Jeney, S., Lukić, B., Kraus, J. A., Franosch, T. & Forró, L. Anisotropic memory effects in confined colloidal diffusion. *Phys. Rev. Lett.* **100**, 240604 (2008).
- Perrin, J. Mouvement brownien et réalité moléculaire. *Ann. Chim. Phys.* **18**, 1–114 (1909).
- Perrin, J. *Atoms* Ch. 3–5 (Constable, 1920).
- Hänggi, P. & Marchesoni, F. 100 years of Brownian motion. *Chaos* **15**, 026101 (2005).
- Berg-Sørensen, K. & Flyvbjerg, H. The colour of thermal noise in classical Brownian motion: a feasibility study of direct experimental observation. *N. J. Phys.* **7**, 38 (2005).
- Fritz, J. *et al.* Translating biomolecular recognition into nanomechanics. *Science* **288**, 316–318 (2000).
- Burg, T. P. *et al.* Weighing of biomolecules, single cells and single nanoparticles in fluid. *Nature* **446**, 1066–1069 (2007).
- Braeckmans, K., De Smedt, S. C., Leblans, M., Pauwels, R. & Demeester, J. Encoding microcarriers: present and future technologies. *Nature Rev. Drug Discov.* **1**, 447–456 (2002).
- Craighead, H. Future lab-on-a-chip technologies for interrogating individual molecules. *Nature* **442**, 387–393 (2006).
- Einstein, A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.* **322**, 549–560 (1905).
- Wang, M. C. & Uhlenbeck, G. E. On the theory of the Brownian motion II. *Rev. Mod. Phys.* **17**, 323–342 (1945).
- Vladimirsky, V. & Terletzky, Y. A. Hydrodynamical theory of translational Brownian motion. *Zh. Eksp. Theor. Fiz.* **15**, 258–263 (1945).
- Hinch, E. J. Application of the Langevin equation to fluid suspensions. *J. Fluid Mech.* **72**, 499–511 (1975).
- Pralle, A., Prummer, M., Florin, E.-L., Stelzer, E. H. K. & Horber, J. K. H. Three-dimensional high-resolution particle tracking for optical tweezers by forward scattered light. *Microsc. Res. Tech.* **44**, 378–386 (1999).
- Jeney, S., Mor, F., Kőszali, R., Forró, L. & Moy, V. T. Monitoring ligand-receptor interactions by photonic force microscopy. *Nanotechnology* **21**, 255102 (2010).
- Lukić, B. *et al.* Motion of a colloidal particle in an optical trap. *Phys. Rev. E* **76**, 011112 (2007).
- Berg-Sørensen, K. & Flyvbjerg, H. Power spectrum analysis for optical tweezers. *Rev. Sci. Instrum.* **75**, 594–612 (2004).
- Clercx, H. J. H. & Schram, P. P. J. M. Brownian particles in shear flow and harmonic potentials: a study of long-time tails. *Phys. Rev. A* **46**, 1942–1950 (1992).
- Huang, R. *et al.* Direct observation of the full transition from ballistic to diffusive Brownian motion in a liquid. *Nature Phys.* **7**, 576–580 (2011).
- Di Leonardo, R. *et al.* Parametric resonance of optically trapped aerosols. *Phys. Rev. Lett.* **99**, 010601 (2007).
- Bormuth, V. *et al.* Optical trapping of coated microspheres. *Opt. Express* **16**, 13831–13844 (2008).
- Gompper, G., Ihle, T., Kroll, D. M. & Winkler, R. G. Advanced computer simulation approaches for soft matter sciences III. *Adv. Polym. Sci.* **221**, 1–87 (2009).
- Zerbe, C., Jung, P. & Hänggi, P. Brownian parametric oscillators. *Phys. Rev. E* **49**, 3626–3635 (1994).
- Pedersen, L. & Flyvbjerg, H. Comment on “Direct measurement of the oscillation frequency in an optical-tweezers trap by parametric excitation”. *Phys. Rev. Lett.* **98**, 189801 (2007).
- Ashkin, A., Dziedzic, J. M. & Yamane, T. Optical trapping and manipulation of single cells using infrared laser beams. *Nature* **330**, 769–771 (1987).
- Peterman, E. J. G., van Dijk, M. A., Kapitein, L. C. & Schmidt, C. F. Extending the bandwidth of optical-tweezers interferometry. *Rev. Sci. Instrum.* **74**, 3246–3249 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements S.J. acknowledges the Swiss National Science Foundation (SNF; grant nos 200021-113529 and 206021-121396). M.G. is supported by NCCR Nanoscale Science and the German Academic Exchange Service (DAAD) and F.M.M. is supported by the National Competence Center in Biomedical Imaging (NCCBI). M.B. and G.F. acknowledge support from the SNF (grant no. PP0022_119006). We thank W. Öffner and R. Kőszali for technical help and U. Aebi, B. U. Felderhof, H. Flyvbjerg, S. Melchionna, E. Sackmann and R. G. Winkler for discussions.

Author Contributions T.F. and M.G. contributed to the planning of the experiments, designed parts of the data analysis software, derived the theory, fitted the theory to the data and interpreted the data. M.B. performed the simulations, analysed the numerical results and contributed to the fitting of the data. F.M.M. contributed to the optimization of the experimental set-up. M.B. and G.F. devised, implemented and tested the numerical simulations. L.F. contributed to the planning of the experiments. S.J. constructed and characterized the experimental set-up; designed, planned and carried out the experiments; designed the data analysis software; and interpreted the data. All authors contributed to, discussed and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.J. (sylvia.jeney@epfl.ch).

Laser cooling of a nanomechanical oscillator into its quantum ground state

Jasper Chan¹, T. P. Mayer Alegre^{1†}, Amir H. Safavi-Naeini¹, Jeff T. Hill¹, Alex Krause¹, Simon Gröblacher^{1,2}, Markus Aspelmeyer² & Oskar Painter¹

The simple mechanical oscillator, canonically consisting of a coupled mass–spring system, is used in a wide variety of sensitive measurements, including the detection of weak forces¹ and small masses². On the one hand, a classical oscillator has a well-defined amplitude of motion; a quantum oscillator, on the other hand, has a lowest-energy state, or ground state, with a finite-amplitude uncertainty corresponding to zero-point motion. On the macroscopic scale of our everyday experience, owing to interactions with its highly fluctuating thermal environment a mechanical oscillator is filled with many energy quanta and its quantum nature is all but hidden. Recently, in experiments performed at temperatures of a few hundredths of a kelvin, engineered nanomechanical resonators coupled to electrical circuits have been measured to be oscillating in their quantum ground state^{3,4}. These experiments, in addition to providing a glimpse into the underlying quantum behaviour of mesoscopic systems consisting of billions of atoms, represent the initial steps towards the use of mechanical devices as tools for quantum metrology^{5,6} or as a means of coupling hybrid quantum systems^{7–9}. Here we report the development of a coupled, nanoscale optical and mechanical resonator¹⁰ formed in a silicon microchip, in which radiation pressure from a laser is used to cool the mechanical motion down to its quantum ground state (reaching an average phonon occupancy number of 0.85 ± 0.08). This cooling is realized at an environmental temperature of 20 K, roughly one thousand times larger than in previous experiments and paves the way for optical control of mesoscale mechanical oscillators in the quantum regime.

It has been known for some time¹¹ that atoms and ions nearly resonant with an applied laser beam (or series of beams) may be mechanically manipulated—even trapped and cooled down to the quantum ground state of their centre-of-mass motion¹². Equally well known¹ is the fact that radiation pressure can be exerted on ordinary (that is, non-resonant) dielectric objects to damp and cool their mechanical motion. In ‘cavity-assisted’ schemes, the radiation pressure force is enhanced by coupling the motion of a mechanical object to the electromagnetic field in a resonant cavity. Pumping of the cavity by a single-frequency electromagnetic source produces a coupling between the mechanical motion and the intensity of the electromagnetic field built up in the resonator. Because the radiation pressure force exerted on the mechanical object is proportional to the field intensity in the resonator, a form of dynamical back-action results^{1,13}. For a lower-frequency (‘red’) detuning of the pump source from the cavity, this leads to damping and cooling of the mechanical motion.

Recent experiments involving micro- and nanomechanical resonators coupled to electromagnetic fields at optical and microwave frequencies have demonstrated significant dynamic back-action due to radiation pressure¹³. These structures have included Fabry–Pérot cavities with mechanically compliant miniature end mirrors^{14–18} or internal nanomembranes¹⁹, whispering-gallery glass resonators²⁰, nanowires capacitively coupled to co-planar microwave transmission line cavities^{6,21} and lumped-circuit microwave resonators with deformable,

nanoscale, vacuum-gap capacitors²². The first measurement of an engineered mesoscopic mechanical resonator predominantly in its quantum ground state, however, was performed not using back-action cooling but rather using conventional cryogenic cooling (bath temperature, $T_b \approx 25$ mK) of a high-frequency and, thus, low-thermal-occupancy oscillator³. Read-out and control of mechanical motion at the single-quantum level was performed by strongly coupling the gigahertz-frequency piezoelectric mechanical resonator to a resonant superconducting quantum circuit. Only recently have microwave systems, also operating at bath temperatures of $T_b \approx 25$ mK, used radiation pressure back-action to cool a high-Q-factor, megahertz-frequency mechanical oscillator to the ground state^{4,21}.

Optically coupled mechanical devices, although they allow for control of the mechanical system through well-established quantum optical techniques²³, have thus far not reached the quantum regime owing to a great number of technical difficulties²⁰. A particular challenge has been maintaining efficient optical coupling and low-loss optics and mechanics in a cryogenic, subkelvin environment. The optomechanical system studied in this work allows large optical coupling to a high-Q, gigahertz-frequency mechanical oscillator, offering both efficient back-action cooling and significantly higher operating temperatures. As shown in Fig. 1a, the system consists of an integrated optical and mechanical nanoscale resonator formed in the surface layer of a silicon-on-insulator microchip. The periodic patterning of the nanobeam is designed to result in Bragg scattering of both optical and acoustic guided waves. A perturbation in the periodicity at the centre of the beam results in co-localized optical and mechanical resonances (Fig. 1b, c), which are coupled through radiation pressure¹⁰. The fundamental optical resonance of the structure occurs at a frequency of $\omega_o/2\pi = 195$ THz ($\lambda = 1,537$ nm), whereas, owing to the speed of sound being much less than the speed of light, the mechanical resonance occurs at $\omega_m/2\pi = 3.68$ GHz. To minimize mechanical damping in the structure, an external acoustic radiation shield is added in the periphery of the nanobeam (Fig. 1d, e). This shield consists of a two-dimensional ‘cross’ pattern, which has been shown both theoretically and experimentally to yield a substantial phononic bandgap in the gigahertz frequency band²⁴.

We use a fibre-taper nanoprobe, formed from standard single-mode optical fibre, to optically couple to the silicon nanoscale resonators. As shown in Fig. 2, a tunable laser (New Focus Velocity swept laser; 200-kHz linewidth) is used to cool optically and transduce the mechanical motion of the nanomechanical oscillator. Placing the optomechanical devices into a continuous-flow helium cryostat provides pre-cooling down to $T_b \approx 20$ K, reducing the bath occupancy of the 3.68-GHz mechanical mode to $n_b \approx 100$. At this temperature, the mechanical Q-factor increases up to a measured value of $Q_m \approx 10^5$, corresponding to an intrinsic mechanical damping rate of $\gamma_i/2\pi = 35$ kHz. The optical Q-factor is measured to be $Q_o = 4 \times 10^5$, corresponding to an optical linewidth of $\kappa/2\pi = 500$ MHz, slightly reduced from its room-temperature value.

¹Thomas J. Watson, Sr, Laboratory of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA. ²Vienna Center for Quantum Science and Technology (VCQ), Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Vienna, Austria. [†]Present address: Instituto de Física “Gleb Wataghin”, Universidade Estadual de Campinas, UNICAMP, 13083-859, Campinas, SP, Brazil.

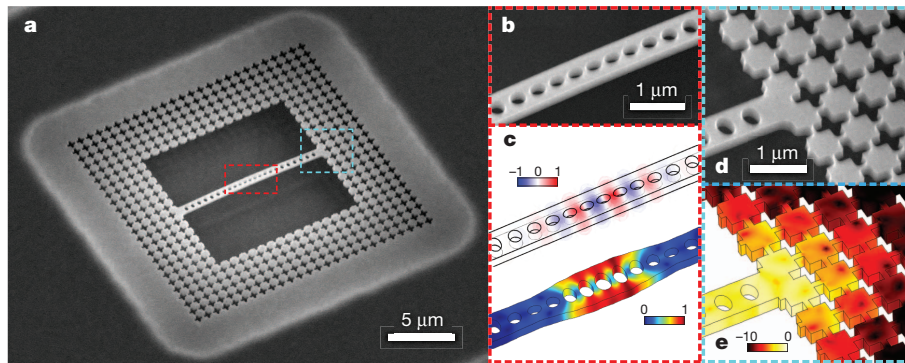


Figure 1 | Optomechanical resonator with phononic shield. **a**, Scanning electron microscope (SEM) image of the patterned silicon nanobeam and the external phononic bandgap shield. **b**, Enlarged SEM image of the central cavity region of the nanobeam. **c**, Top: normalized electric field (colour scale) of the localized optical resonance of the nanobeam cavity, simulated using the finite-element method (FEM). Bottom: FEM simulation of the normalized displacement field of the acoustic resonance (breathing mode), which is coupled by radiation pressure to the co-localized optical resonance. The

displacement field is indicated by the exaggerated deformation of the structure, with the relative magnitude of the local displacement (strain) indicated by the colour. **d**, SEM image of the interface between the nanobeam and the phononic bandgap shield. **e**, FEM simulation of the normalized squared displacement field amplitude of the localized acoustic resonance at the nanobeam-shield interface, indicating the strong suppression of acoustic radiation provided by the phononic bandgap shield. The colour scale represents $\log[x^2/\max(x^2)]$, where x is the displacement field amplitude.

In the resolved-sideband limit, where $\omega_m/\kappa > 1$, driving the system with a laser (frequency, ω_l) tuned to the red side of the optical cavity (detuning, $\Delta \equiv \omega_o - \omega_l = \omega_m$), creates an optically induced damping, γ_{OM} , of the mechanical resonance²⁵. In the weak-coupling regime ($\gamma_{OM} \ll \kappa$), the optical back-action damping is given by $\gamma_{OM} = 4g^2 n_c / \kappa$, where n_c is the average number of drive-laser photons stored in the cavity and g is the optomechanical coupling rate between the mechanical and optical modes. This coupling rate, g , is quantified as the shift in the optical resonance for an amplitude of motion equal to

the zero-point fluctuation amplitude ($x_{zpf} = (\hbar/2m\omega_m)^{1/2}$, where m is the motional mass of the localized acoustic mode and \hbar is Planck's constant divided by 2π). The optomechanical damping, which is a result of the preferential scattering of drive photons into the upper-frequency sideband, also cools the mechanical mode. For a quantum-limited drive laser, the phonon occupancy of the mechanical oscillator can be reduced from $n_b = k_B T_b / \hbar \omega_m \gg 1$ to $\bar{n} = n_b / (1 + C) + n_{\min}$, where k_B is Boltzmann's constant and $C \equiv \gamma_{OM} / \gamma_i$ is the cooperativity. The residual scattering of drive photons into the lower-frequency sideband limits the cooled phonon occupancy to $n_{\min} = (\kappa/4\omega_m)^2$, which is determined by the level of sideband resolution²⁵.

The drive laser, in addition to providing mechanical damping and cooling, can be used to measure the mechanical and optical properties of the system through a series of calibrated measurements. In a first set of measurements, we use the noise power spectral density (PSD) of the drive laser transmitted through the optomechanical cavity to perform spectroscopy of the mechanical mode. As shown in Supplementary Information, the noise PSD of the photocurrent generated by the transmitted field of the drive laser with red-sideband detuning ($\Delta = \omega_m$) yields a Lorentzian component of the single-sided PSD proportional to $S_b(\omega) = \bar{n}\gamma / ((\omega - \omega_m)^2 + (\gamma/2)^2)$, where $\gamma = \gamma_i + \gamma_{OM} = \gamma_i(1 + C)$ is the total mechanical damping rate. For a blue laser detuning of $\Delta = -\omega_m$, the optically induced damping is negative ($\gamma_{OM} = -4g^2 n_c / \kappa$) and the photocurrent noise PSD is proportional to $S_{b^*}(\omega) = (\bar{n} + 1)\gamma / ((\omega - \omega_m)^2 + (\gamma/2)^2)$. Typical measured noise power spectra under low-power laser drive ($n_c = 1.4$, $C = 0.27$), for both red and blue detuning, are shown in Fig. 3a. Even at these small drive powers, the effects of back-action on the measured spectra are evident, with the red-detuned drive broadening the mechanical line and the blue-detuned drive narrowing the line. The noise floor in Fig. 3a (shaded in grey) corresponds to the noise generated by the EDFA used to pre-amplify the transmitted drive-laser signal before photodetection, and is several orders of magnitude greater than the electronic noise of the photoreceiver and the real-time spectrum analyser.

Calibration of the EDFA gain, along with the photoreceiver and real-time spectrum analyser photodetection gain, makes it possible to convert the measured area under the photocurrent noise PSD into a mechanical mode phonon occupancy. As described in detail in Supplementary Information, we perform these calibrations, along with measurements of low-drive-power ($C \ll 1$), radio-frequency spectra of both detunings ($\Delta = \pm \omega_m$), to provide accurate, local thermometry of the optomechanical cavity. An example of this form of calibrated mode thermometry is shown in Fig. 3b, where we plot the optically measured

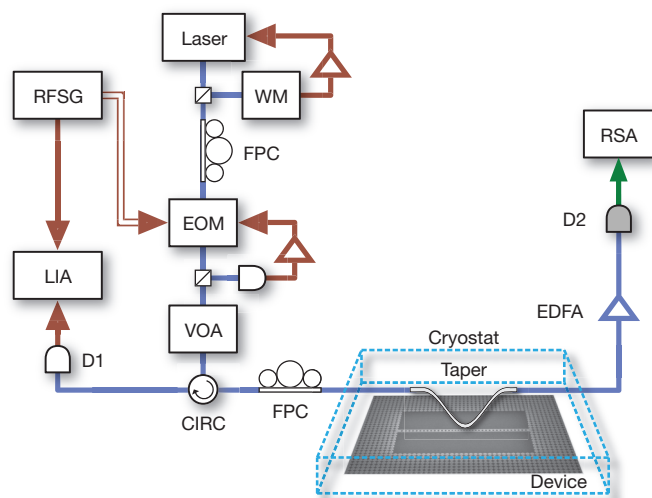


Figure 2 | Experimental set-up. A single, tunable, 1,550-nm diode laser is used as the cooling and mechanical transduction beam sent into the nanobeam optomechanical resonator cavity held in a continuous-flow helium cryostat. A wavemeter (WM) is used to track and lock the laser frequency, and a variable optical attenuator (VOA) is used to set the laser power. The transmitted signal is amplified by an erbium-doped fibre amplifier (EDFA) and detected on a high-speed photodetector (D2) connected to a real-time spectrum analyser (RSA), where the mechanical noise power spectrum is measured. A slowly modulated probe signal used for optical spectroscopy and calibration is generated from the cooling laser beam using an amplitude electro-optic modulator (EOM) driven by a microwave source (RFSG). The reflected component of this signal is separated from the input by an optical circulator (CIRC), sent to a photodetector (D1) and then demodulated using a lock-in amplifier (LIA). Paddle-wheel fibre polarization controllers (FPCs) are used to set the laser polarization at the input to the EOM and the input to the optomechanical cavity. For more detail, see Supplementary Information.

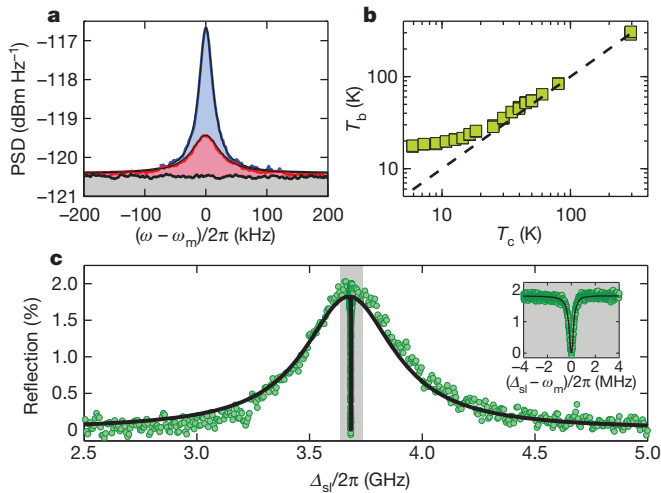


Figure 3 | Mechanical and optical response. **a**, Typical measured mechanical noise spectra around the resonance frequency of the breathing mode for low drive-laser power ($n_c = 1.4$). The blue and red curves correspond to the spectra measured with the drive laser blue- and, respectively, red-detuned by a mechanical frequency from the optical cavity resonance. The black trace corresponds to the measured noise floor (dominated by EDFA noise) with the drive laser detuned far from the cavity resonance. **b**, Plot of the measured (squares) mechanical mode bath temperature (T_b) as a function of cryostat sample mount temperature (T_c). The dashed line indicates the curve corresponding to perfect following of the cryostat temperature by the mode temperature ($T_b = T_c$). **c**, Typical reflection spectrum (normalized power reflection) of the cavity while driven by the cooling laser ($\Delta = \omega_m$, $n_c = 56$, $C = 11$), as measured by a weaker probe beam at two-photon detuning Δ_{sl} . The signature reflection dip on resonance with the bare cavity mode, highlighted in the inset, is indicative of EIT caused by coupling of the optical and mechanical degrees of freedom by the cooling laser beam.

mechanical mode bath temperature, T_b , as a function of the cryostat sample mount temperature, T_c (independently measured using a silicon diode thermometer attached to the copper sample mount).

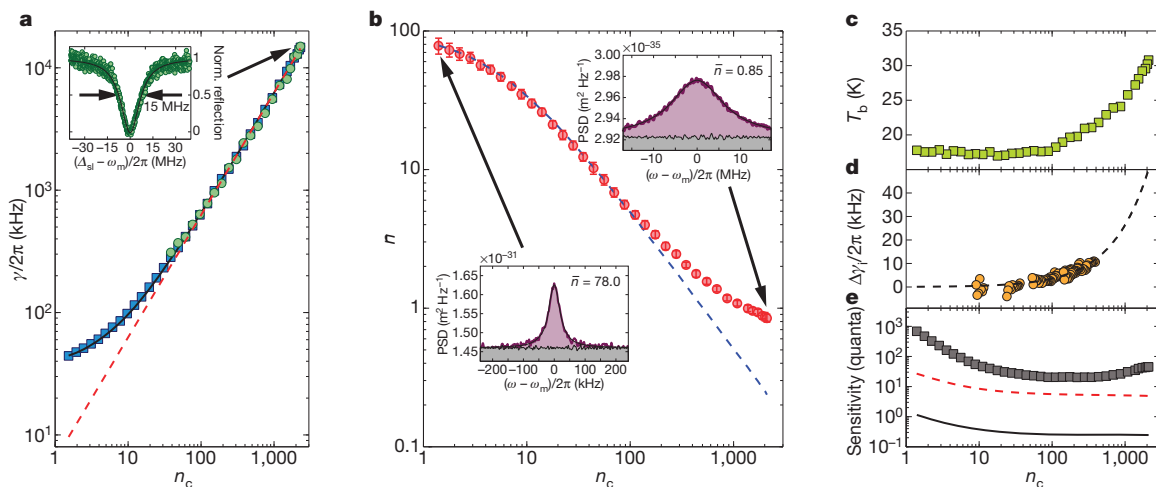


Figure 4 | Optical cooling results. **a**, Measured mechanical mode linewidth (squares), EIT transparency bandwidth (circles) and predicted optomechanical damping rate estimated using the zero-point optomechanical coupling rate, $g/2\pi = 910$ kHz (red dashed line). Inset, measured EIT transparency window at the highest cooling-beam drive power. **b**, Measured (circles) average phonon number, \bar{n} , in the breathing mechanical mode at $\omega_m/2\pi = 3.68$ GHz, versus cooling drive-laser power (in units of intracavity photons, n_c), as deduced from the calibrated area under the Lorentzian line shape of the mechanical noise power spectrum. The inset spectra show the measured noise PSD (using $x_{zpf} = 2.7$ fm, corresponding to the numerically computed motional mass for the breathing mode with $m = 311$ fg). The dashed blue line indicates the estimated mode phonon number calculated from the measured optical

Figure 3b shows that the optical mode thermometry predicts a mode temperature in good correspondence with the absolute temperature of the sample mount for $T_c > 50$ K; below this value, the mode temperature deviates from T_c and saturates to a value of $T_b = 17.6 \pm 0.8$ K owing to thermal radiative heating of the device through the imaging aperture in the radiation shield of our cryostat.

In a second set of measurements, we determine the mechanical damping, γ , and the cavity–laser detuning, Δ , by optical spectroscopy of the driven cavity. By sweeping a second probe beam, of frequency ω_s , over the cavity, with the cooling beam tuned to $\Delta = \omega_m$, spectra showing electromagnetically induced transparency²⁶ (EIT) are measured (Fig. 3c). Owing to the high single-photon cooperativity of the system, an intracavity population of only $n_c \approx 5$ switches the system from reflecting to transmitting for the probe beam. The corresponding dip at the centre of the optical cavity resonance occurs at a two-photon detuning of $\Delta_{sl} \equiv \omega_s - \omega_l = \omega_m$ and has a bandwidth equal to the mechanical damping rate, $\gamma_i(1 + C)$. In Fig. 4a, we plot the measured mechanical linewidth as a function of intracavity photon number, showing good correspondence between both mechanical and optical spectroscopy techniques, and indicating that the system remains in the weak-coupling regime for all measured cooling powers. From a fit to the measured mechanical damping rate as a function of n_c (Fig. 4a, dashed red line), the zero-point optomechanical coupling rate is determined to be $g/2\pi = 910$ kHz.

In Fig. 4b, we plot the calibrated Lorentzian noise PSD area, in units of phonon occupancy, as a function of red-detuned ($\Delta = \omega_m$) drive-laser power. Owing to the low effective temperature of the laser drive, the mechanical mode is not only damped but is also cooled substantially. The minimum measured average mode occupancy for the highest drive power (corresponding to $n_c \approx 2,000$) is $\bar{n} = 0.85 \pm 0.08$, putting the mechanical oscillator in a thermal state with ground-state occupancy probability greater than 50%. The dashed blue line in Fig. 4b represents the ideal back-action-cooled phonon occupancy estimated using both the measured mechanical damping rate in Fig. 4a and the low-drive-power intrinsic mechanical damping rate. Deviation of the measured phonon occupancy from the ideal cooling model is seen to

damping alone. Error bars indicate estimated uncertainties as outlined in Supplementary Information. **c**, Estimated bath temperature, T_b , versus cooling laser intracavity photon number, n_c . **d**, Measured change in the intrinsic mechanical damping rate versus n_c (circles). A polynomial fit to the mechanical damping dependence on n_c is shown as a dashed line. For more details, see Supplementary Information. **e**, The measured (squares) background noise PSD versus drive-laser power (n_c), in units of effective phonon quanta. The red dashed curve corresponds to the theoretical imprecision assuming shot-noise-limited detection but all other cavity properties and optical loss as in the experiment. The solid black curve is for an ideal, quantum-limited continuous position measurement of mechanical motion.

occur at the highest drive powers and results from both an increase in the bath temperature due to optical absorption (Fig. 4c) and an increase in the intrinsic mechanical damping rate (Fig. 4d) induced by the generation of free carriers through optical absorption (Supplementary Information). To evaluate the efficiency of the optical transduction of the mechanical motion, we also plot (Fig. 4e) the measured background noise PSD, or imprecision level. The minimum measured imprecision, occurring for $n_c \approx 500$ corresponds to $n_{\text{imp}} \approx 20$ in units of phonon quanta when referred to the peak Lorentzian level of the transduced mechanical motion (Supplementary Information). Comparing the measured imprecision with the respective theoretical imprecision levels for shot-noise-limited detection (Fig. 4e, dashed red curve) and ideal quantum-limited motion transduction (Fig. 4e, black curve) indicates that $n_{\text{imp}}^{\text{EDFA}} \approx 15$ stems from the excess noise imparted by the EDFA optical amplifier. The remaining $n_{\text{imp}}^{\text{loss}} \approx 5$ is due to optical loss of signal inside the cavity (11.7 dB) and in the optical fibre output waveguide (2 dB).

Looking forward, the optical back-action cooling and thermometry, as performed in this work, represents only a first step towards optical measurement and control of the quantum state of a nanomechanical object. The mechanical system, although cooled to a mode occupancy of less than one, is still prepared in a classical thermal state, with its quantum zero-point fluctuations hidden by our measurement scheme. However, experiments to prepare and measure non-classical quantum states of the mechanical system are now within reach. A basic requirement for optomechanical experiments in the quantum regime is the ability to exchange photons with the mechanical resonator on a timescale shorter than that for a single thermal phonon to enter the mechanical system from the environment. The latter, called the thermal decoherence time, is given by $\tau_{\text{th}} \equiv \hbar Q_{\text{m}} / k_{\text{B}} T_{\text{b}}$, and the timescale on which the mechanical resonator exchanges photons with an optical input is $\tau_{\text{OM}} \equiv 1/\gamma_{\text{OM}}$. The requirement that $\tau_{\text{OM}} < \tau_{\text{th}}$ is equivalent to the requirement for optical back-action cooling of the mechanical oscillator to $\bar{n} < 1$, and is thus realized for the optomechanical crystal devices reported here. This allows for optomechanical entanglement between light and mechanics²⁷ or quantum state transfer between single optical photons and mechanical phonons^{9,28}, enabling mechanical systems to function as both quantum transducers⁸ and quantum memory elements²⁹. In addition, the chip-scale nature of the optomechanical crystal architecture naturally lends itself to the creation of coupled photon-phonon circuits, facilitating not only the coupling of multiple mechanical and optical objects, but also allowing for the integration of optomechanics with other quantum systems such as superconducting quantum circuits⁹. Finally, if a regime of strong coupling at the single-quantum level³⁰ ($g/\kappa > 1$) could be reached, many new opportunities would be available, not least the study of nonlinear phononics at the single-phonon level and the generation of highly non-classical quantum states in mechanical or optical systems.

Received 17 June; accepted 16 August 2011.

1. Braginsky, V. & Manukin, A. *Measurement of Weak Forces in Physics Experiments* (Univ. Chicago Press, 1977).
2. Jensen, K., Kim, K. & Zettl, A. An atomic-resolution nanomechanical mass sensor. *Nature Nanotechnol.* **3**, 533–537 (2008).
3. O'Connell, A. D. *et al.* Quantum ground state and single-phonon control of a mechanical resonator. *Nature* **464**, 697–703 (2010).
4. Teufel, J. D. *et al.* Sideband cooling of micromechanical motion to the quantum ground state. *Nature* **475**, 359–363 (2011).
5. Caves, C., Thorne, K., Drever, R., Sandberg, V. D. & Zimmermann, M. On the measurement of a weak classical force coupled to a quantum-mechanical oscillator. *Rev. Mod. Phys.* **52**, 341–392 (1980).

6. Regal, C. A., Teufel, J. D. & Lehnert, K. W. Measuring nanomechanical motion with a microwave cavity interferometer. *Nature Phys.* **4**, 555–560 (2008).
7. Wallquist, M., Hammerer, K., Rabl, P., Lukin, M. & Zoller, P. Hybrid quantum devices and quantum engineering. *Phys. Scr.* **2009**, 014001 (2009).
8. Stannigel, K., Rabl, P., Sørensen, A. S., Zoller, P. & Lukin, M. D. Optomechanical transducers for long-distance quantum communication. *Phys. Rev. Lett.* **105**, 220501 (2010).
9. Safavi-Naeini, A. H. & Painter, O. Proposal for an optomechanical traveling wave phonon-photon translator. *N. J. Phys.* **13**, 013017 (2011).
10. Eichenfield, M., Chan, J., Camacho, R. M., Vahala, K. J. & Painter, O. Optomechanical crystals. *Nature* **462**, 78–82 (2009).
11. Cohen-Tannoudji, C. N. & Phillips, W. D. New mechanisms for laser cooling. *Phys. Today* **43**, 33–40 (1990).
12. Diedrich, F., Bergquist, J. C., Itano, W. M. & Wineland, D. J. Laser cooling to the zero-point energy of motion. *Phys. Rev. Lett.* **62**, 403–406 (1989).
13. Kippenberg, T. J. & Vahala, K. J. Cavity optomechanics: back-action at the mesoscale. *Science* **321**, 1172–1176 (2008).
14. Cohadon, P.-F., Heidmann, A. & Pinard, M. Cooling of a mirror by radiation pressure. *Phys. Rev. Lett.* **83**, 3174–3177 (1999).
15. Metzger, C. H. & Karrai, K. Cavity cooling of a microlever. *Nature* **432**, 1002–1005 (2004).
16. Gigan, S. *et al.* Self-cooling of a micromirror by radiation pressure. *Nature* **444**, 67–70 (2006).
17. Arcizet, O., Cohadon, P.-F., Briant, T., Pinard, M. & Heidmann, A. Radiation-pressure cooling and optomechanical instability of a micromirror. *Nature* **444**, 71–74 (2006).
18. Gröblacher, S. *et al.* Demonstration of an ultracold micro-optomechanical oscillator in a cryogenic cavity. *Nature Phys.* **5**, 485–488 (2009).
19. Thompson, J. D. *et al.* Strong dispersive coupling of a high-finesse cavity to a micromechanical membrane. *Nature* **452**, 72–75 (2008).
20. Rivière, R. *et al.* Optomechanical sideband cooling of a micromechanical oscillator close to the quantum ground state. *Phys. Rev. A* **83**, 063835 (2011).
21. Rocheleau, T. *et al.* Preparation and detection of a mechanical resonator near the ground state of motion. *Nature* **463**, 72–75 (2010).
22. Teufel, J. D. *et al.* Circuit cavity electromechanics in the strong-coupling regime. *Nature* **471**, 204–208 (2011).
23. Wiseman, H. M. & Milburn, G. J. *Quantum Measurement and Control* (Cambridge Univ. Press, 2010).
24. Alegre, T. P. M., Safavi-Naeini, A., Winger, M. & Painter, O. Quasi-two-dimensional optomechanical crystals with a complete phononic bandgap. *Opt. Express* **19**, 5658–5669 (2011).
25. Marquardt, F., Chen, J. P., Clerk, A. A. & Girvin, S. M. Quantum theory of cavity-assisted sideband cooling of mechanical motion. *Phys. Rev. Lett.* **99**, 093902 (2007).
26. Safavi-Naeini, A. H. *et al.* Electromagnetically induced transparency and slow light with optomechanics. *Nature* **472**, 69–73 (2011).
27. Vitali, D. *et al.* Optomechanical entanglement between a movable mirror and a cavity field. *Phys. Rev. Lett.* **98**, 030405 (2007).
28. Akram, U., Kiesel, N., Aspelmeyer, M. & Milburn, G. J. Single-photon optomechanics in the strong coupling regime. *N. J. Phys.* **12**, 083030 (2010).
29. Chang, D., Safavi-Naeini, A. H., Hafezi, M. & Painter, O. Slowing and stopping light using an optomechanical crystal array. *N. J. Phys.* **13**, 023003 (2011).
30. Rabl, P. Photon blockade effect in optomechanical systems. *Phys. Rev. Lett.* **107**, 063601 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the DARPA/MTO ORCHID program through a grant from the AFOSR, the European Commission (MINOS, QUESSANCE), the European Research Council (ERC QOM), the Austrian Science Fund (CoQuS, FOQUS, START) and the Kavli Nanoscience Institute at the California Institute of Technology. The authors thank B. Baker for help with the cryostat set-up, J.C. thanks R. Li, and J.C. and A.H.S.-N. acknowledge support from NSERC.

Author Contributions J.C., T.P.M.A. and A.H.S.-N. designed the device, and J.C. fabricated it with support from J.T.H. J.C., T.P.M.A., A.H.S.-N., J.T.H., A.K. and S.G. performed the measurements and analysed the measured data. O.P. and M.A. supervised the measurements and the data analysis. All authors contributed to the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to O.P. (opainter@caltech.edu).

A carbon isotope challenge to the snowball Earth

P. Sansjofre^{1,2}, M. Ader¹, R. I. F. Trindade², M. Elie^{3†}, J. Lyons⁴, P. Cartigny¹ & A. C. R. Nogueira⁵

The snowball Earth hypothesis postulates that the planet was entirely covered by ice for millions of years in the Neoproterozoic era, in a self-enhanced glaciation caused by the high albedo of the ice-covered planet. In a hard-snowball picture, the subsequent rapid unfreezing resulted from an ultra-greenhouse event attributed to the buildup of volcanic carbon dioxide (CO_2) during glaciation¹. High partial pressures of atmospheric CO_2 (p_{CO_2} ; from 20,000 to 90,000 p.p.m.v.) in the aftermath of the Marinoan glaciation (~635 Myr ago) have been inferred from both boron and triple oxygen isotopes^{2,3}. These p_{CO_2} values are 50 to 225 times higher than present-day levels. Here, we re-evaluate these estimates using paired carbon isotopic data for carbonate layers that cap Neoproterozoic glacial deposits and are considered to record post-glacial sea level rise¹. The new data reported here for Brazilian cap carbonates, together with previous ones for time-equivalent units^{4–8}, provide p_{CO_2} estimates lower than 3,200 p.p.m.v.—and possibly as low as the current value of ~400 p.p.m.v. Our new constraint, and our re-interpretation of the boron and triple oxygen isotope data, provide a completely different picture of the late Neoproterozoic environment, with low atmospheric concentrations of carbon dioxide and oxygen that are inconsistent with a hard-snowball Earth.

Thousands of carbon isotope data have been reported for Neoproterozoic successions in the past decade^{4–8}, yet the full palaeoenvironmental significance of these data is still largely unappreciated. In particular, coupled carbon isotope data from organic carbon and carbonate have the potential to solve the longstanding conundrum of carbon dioxide concentrations in the aftermath of Neoproterozoic glaciations. This is possible because the difference between the carbon isotope ratio for carbonates ($\delta^{13}\text{C}_{\text{carb}}$) and that for associated organic matter ($\delta^{13}\text{C}_{\text{org}}$, $\Delta^{13}\text{C}_{\text{carb-org}}$) depends strongly on the concentration of dissolved CO_2 in the ocean ($[\text{CO}_2]_{\text{aq}}$; ref. 9), which can be related to p_{CO_2} (ref. 10). This is illustrated by the decrease of $\Delta^{13}\text{C}_{\text{carb-org}}$ in the past 20 Myr to today's value of ~22‰ (ref. 9), which accompanied the drawdown of atmospheric p_{CO_2} to the pre-industrial value of 280 p.p.m.v.. In contrast, earlier in the Phanerozoic eon, $\Delta^{13}\text{C}_{\text{carb-org}}$ mostly remained in the range 28–32‰ (ref. 9), except for brief episodes of lower $\Delta^{13}\text{C}_{\text{carb-org}}$, such as those reported in the upper Ordovician¹¹ and at the Permo-Triassic boundary¹².

We obtained paired $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{org}}$ values for post-glacial cap carbonates from western Brazil. Cap carbonates are the stratigraphic horizon marker defining the base of the Ediacaran period (635–542 Myr before present). They were supposedly deposited during the post-glacial sea level rise, in a supersaturated ocean and ultra-greenhouse climate resulting from high atmospheric CO_2 levels following a snowball Earth¹. The studied cap carbonates come from the southeastern margin of the Amazonian craton, away from the metamorphic Paraguay belt¹³. They form a transgressive systems tract directly above diamictites of the Puga Formation, starting with pink stromatolitic dolostones of the Mirassol d'Oeste Formation (~13 m thick) deposited in shallow oxic waters, overlain by dark grey limestone and shale of the Guia Formation (~50 m thick) deposited below storm wave base. These strata have been correlated to the Marinoan

event on the basis of their $\delta^{13}\text{C}_{\text{carb}}$ of about –5‰, $^{87}\text{Sr}/^{86}\text{Sr}$ of ~0.7078 and Pb–Pb carbonate age of 627 ± 32 Myr (Fig. 1a; Supplementary Information section 2). The measured $\delta^{13}\text{C}_{\text{org}}$ values ($n = 29$) are homogeneous, with an average value of -27.3 ± 0.5 ‰ (Supplementary Table 1). $\delta^{13}\text{C}_{\text{carb}}$ values are also very homogeneous, averaging -4.8 ± 0.6 ‰, consistent with previous results on the same unit and on other cap carbonates worldwide (Supplementary Fig. 8). The resulting $\Delta^{13}\text{C}_{\text{carb-org}}$ values average to 22.7 ± 0.8 ‰ (Fig. 1a).

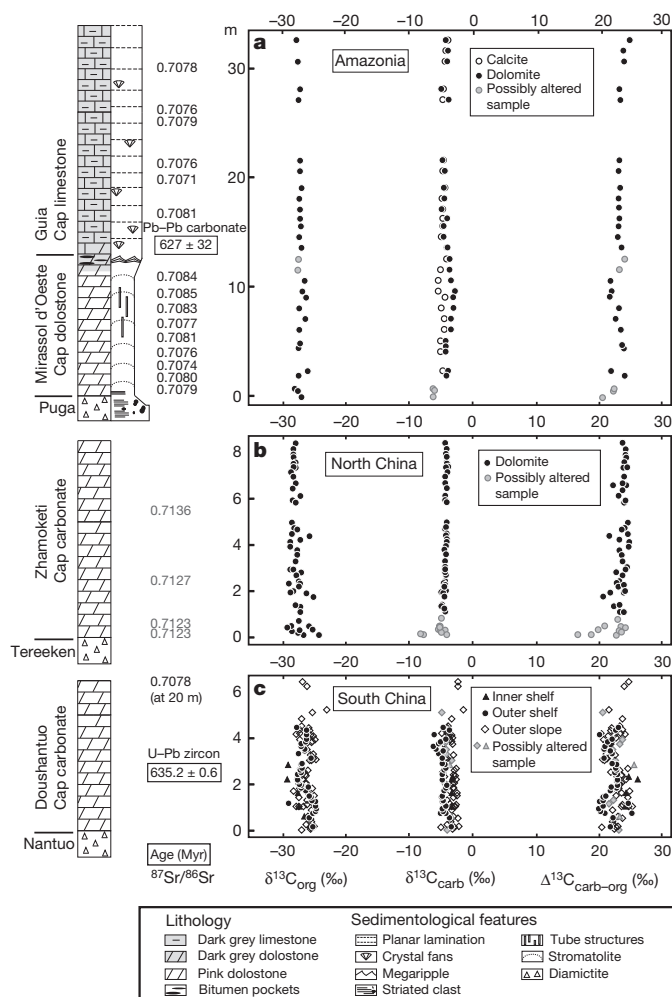


Figure 1 | Isotope and age data for cap carbonates. Paired carbon isotope data ($\delta^{13}\text{C}_{\text{org}}$ and $\delta^{13}\text{C}_{\text{carb}}$) for cap carbonate successions from Brazil (Amazonia), North China and South China (this work and refs 4–8), along with geochronological data and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (see Supplementary Information sections 2.1 and 2.3 for references). Potentially altered $^{87}\text{Sr}/^{86}\text{Sr}$ values for Zhamoketi cap dolostones are shown in grey.

¹Équipe de Géochimie des Isotopes Stables, Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Univ. Paris Diderot, UMR 7154 CNRS, 75238 Paris Cedex 05, France. ²Departamento de Geofísica, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Rua do Matão 1226, 05508-900 São Paulo, Brazil. ³UMR CNRS 7566 G2R, Nancy-Université, CNRS, BP 239, 54506 Vandœuvre-lès-Nancy, France. ⁴Institute of Geophysics and Planetary Physics, University of California, Los Angeles, 595 Charles Young Dr. East, Los Angeles, California 90095-1567, USA. ⁵Faculdade de Geologia, Instituto de Geociências, Universidade Federal do Pará, CEP 66.075-110, Belém, Brazil. [†]Present address: Sarawak Shell Berhad Locked Bag No. 1, 98009 Miri, Sarawak, Malaysia.

All other Marinoan cap carbonates for which paired carbon isotope data are available present similarly low $\Delta^{13}\text{C}_{\text{carb-org}}$ values (Fig. 1). In North China, the 10-m-thick Zhamoketi cap carbonate, deposited above the Tereken diamictite⁴, shows a very stable $\Delta^{13}\text{C}_{\text{carb-org}}$ signal of $23.6 \pm 1.5\text{‰}$ ($n = 52$, Fig. 1b). In South China, $\Delta^{13}\text{C}_{\text{carb-org}}$ data for the 3–6-m-thick Doushantuo cap carbonates are available for six sections located along a north–south transect across the Yangtze platform^{5–8}. These sections reveal again a low $\Delta^{13}\text{C}_{\text{carb-org}}$ signal, with an average of $22.7 \pm 1.4\text{‰}$ ($n = 105$, Fig. 1c). Other cap carbonate successions also show systematically low values: the Noonday dolomite, Death Valley (average $19.1 \pm 2.7\text{‰}$, $n = 9$)¹⁴, the Maieberg Formation, Namibia (average $18.7 \pm 0.8\text{‰}$, $n = 8$)¹⁵ and the Tepee dolostone, Western Canada (single value of 22.4‰)¹⁶. These sections are not considered further, owing to their much smaller data sets.

We have evaluated the extent to which post-depositional processes may have overprinted $\delta^{13}\text{C}_{\text{carb}}$ (ref. 17) and $\delta^{13}\text{C}_{\text{org}}$ (ref. 9) in these sections. Available indicators (petrography, Mn/Sr and $\delta^{18}\text{O}_{\text{carb}}$; Supplementary Information sections 3 and 6, respectively) and the smooth chemostratigraphic trends observed in different successions with similar $\delta^{13}\text{C}_{\text{carb}}$ (Fig. 1) argue against a significant diagenetic overprint for the $\delta^{13}\text{C}_{\text{carb}}$ data. As for $\delta^{13}\text{C}_{\text{org}}$, isotope effects associated with early diagenesis are minor and taken into account in the A_2 parameter of equation (1) below (Supplementary Information section 5.1). In addition, molecular organic geochemistry and Rock-Eval pyrolysis data for Brazilian cap carbonates (Supplementary Information section 4) show that the organic matter experienced low thermal maturity and only moderate oxidative weathering, the $\delta^{13}\text{C}_{\text{org}}$ signal being thus not significantly affected by these processes (Supplementary Information section 5). After screening for post-depositional effects, 26 data (out of 186) were eliminated, and a grand mean $\Delta^{13}\text{C}_{\text{carb-org}}$ of $23.2 \pm 0.9\text{‰}$ was calculated for the three cap carbonate sequences (Fig. 1). The strong similarity in paired carbon isotope values among the three sequences constitutes the most compelling argument against a first-order diagenetic control on their low $\Delta^{13}\text{C}_{\text{carb-org}}$.

Anomalous $\Delta^{13}\text{C}_{\text{carb-org}}$ values in Neoproterozoic successions were previously interpreted as the result of organic matter input from a large dissolved organic carbon (DOC) pool in the ocean⁶. A large DOC reservoir can accumulate only in anoxic waters and would buffer the $\delta^{13}\text{C}_{\text{org}}$, thus decoupling the $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{org}}$ signals. However, several authors have recently challenged this hypothesis, on the basis of mass balance calculations^{6,18}. Moreover, to account for the consistently low $\Delta^{13}\text{C}_{\text{carb-org}}$ values observed here, cap carbonates must form exclusively in DOC-rich anoxic waters. This contradicts geochemical and magnetic data showing that the basal haematite-bearing cap dolostones were deposited in oxic surface waters¹⁹.

We thus consider that the cap carbonates and associated organic matter originated from the same surface water, so their $\Delta^{13}\text{C}_{\text{carb-org}}$ can be expressed as⁹

$$\Delta^{13}\text{C}_{\text{carb-org}} = \varepsilon_p - A_2 + A_{\text{carb}} \quad (1)$$

where ε_p is the photosynthetic fractionation factor between dissolved CO_2 and organic matter and depends on $[\text{CO}_2]_{\text{aq}}$; A_2 is the potential increase in $\delta^{13}\text{C}_{\text{org}}$ during early diagenesis, set here at $+1.5\text{‰}$ (see Methods); and A_{carb} is the isotopic depletion of dissolved CO_2 relative to carbonate. A_{carb} is temperature-dependent and can be estimated from the carbon isotope fractionation between carbonate species (Methods).

The range of ‘normal’ ε_p values recorded for most of the Phanerozoic eon⁹ (Fig. 2b) matches the low $\Delta^{13}\text{C}_{\text{carb-org}}$ of cap carbonates ($22.3\text{‰} < \Delta^{13}\text{C}_{\text{carb-org}} < 24.1\text{‰}$) only for temperatures higher than 80°C . However, the activase of the photosynthetic enzyme (Rubisco) is ineffective above 45°C (ref. 20), and only non-phototrophic hyperthermophiles can survive in extremely warm environments²¹. As temperatures above 80°C would be at odds with the presence of

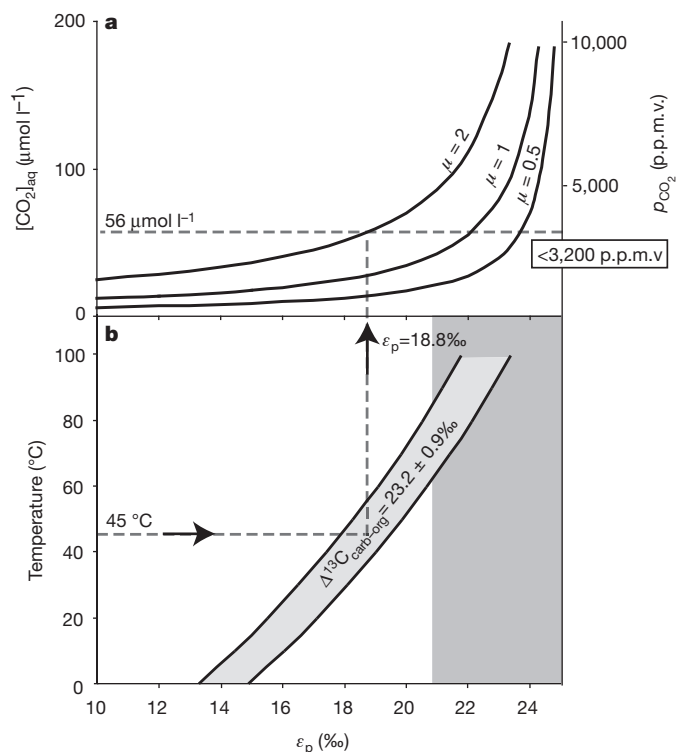


Figure 2 | Relationship between photosynthetic fractionation factor (ε_p), temperature and CO_2 concentrations. **a**, Dissolved CO_2 in the ocean ($[\text{CO}_2]_{\text{aq}}$) and atmospheric $p\text{CO}_2$ versus ε_p for three photosynthesizer growth rates (μ , in d^{-1}), obtained from the relationship $[\text{CO}_2]_{\text{aq}} = 182 \mu(V/S)/(25.3\varepsilon_p)$, and assuming chemical equilibrium between ocean and atmosphere. In this equation²⁷, 25.3‰ is the isotope effect associated with carbon fixation, and the coefficient 182 is partly dependent on cell membrane permeability. **b**, ε_p – T interval compatible with $\Delta^{13}\text{C}_{\text{carb-org}}$ values of cap carbonates ($22.3 < \Delta^{13}\text{C}_{\text{carb-org}} < 24.1\text{‰}$). At 45°C , average $\Delta^{13}\text{C}_{\text{carb-org}}$ implies an ε_p of 18.8‰ , and hence a $[\text{CO}_2]_{\text{aq}}$ of $55.9 \mu\text{mol l}^{-1}$, corresponding to an atmospheric $p\text{CO}_2$ of $3,200$ p.p.m.v.. See text and Methods for details. The shaded area in **b** shows the range of ε_p observed during most of the Phanerozoic eon.

autotroph and heterotroph fossils within and above glacial deposits^{22,23} (Supplementary Information section 4), the low $\Delta^{13}\text{C}_{\text{carb-org}}$ recorded in cap carbonates instead implies low ε_p values (from 18.8‰ at 45°C to 16.9‰ at 25°C), similar to those of today²⁴.

The low ε_p values of cap carbonates are consistent with a low $[\text{CO}_2]_{\text{aq}}$ during their deposition, regardless of the speciation of carbon (CO_2 or HCO_3^-) and of its uptake mechanism. In some modern settings, ε_p is shifted towards lower values, probably because the system is significantly driven by the active uptake of HCO_3^- (refs 25, 26). But controlled experiments show that the relative contribution of this carbon uptake mechanism decreases significantly as $[\text{CO}_2]_{\text{aq}}$ increases²⁵. For instance, it did not influence ε_p for most of the Phanerozoic, when $[\text{CO}_2]_{\text{aq}}$ was probably much higher than today⁹. Therefore, this mechanism could account for the systematically low ε_p deduced here only if $[\text{CO}_2]_{\text{aq}}$ during cap carbonate formation were low—probably below the present-day range of 10 – $20 \mu\text{mol kg}^{-1}$. In the case of dissolved CO_2 uptake, ε_p values can be directly related to $[\text{CO}_2]_{\text{aq}}$ by empirical equations, which include physiological parameters such as the ratio of cellular volume to surface area, (V/S), and the growth rate (μ) of photosynthesizers. We used the $[\text{CO}_2]_{\text{aq}}$ – ε_p equation from ref. 27 for modern unicellular algal species (Fig. 2a); other equations give similar or lower estimates (see Methods). Figure 2a shows three curves of $[\text{CO}_2]_{\text{aq}}$ as a function of ε_p for different growth rates and for a V/S of $1 \mu\text{m}$. Taking a maximum temperature of 45°C and a conservative growth rate of 2d^{-1} , which is rarely attained even in modern high-productivity settings (Supplementary Information section 1), we estimate an absolute $[\text{CO}_2]_{\text{aq}}$ upper limit of $56 \mu\text{mol kg}^{-1}$.

Theoretically, much higher $[\text{CO}_2]_{\text{aq}}$ values could also produce the low ε_p observed in cap carbonates, but only for extremely high cellular volume and V/S ratios. We note, however, that even if large organisms did flourish just after the glaciation, the increase in cell volume of the biomass would probably be counterbalanced by a large decrease in growth rates. Growth rate and cell volume are related by a power-law, $\mu = aV^b$ (a being a normalization factor and b a size-scaling exponent usually taken as -0.25 , ref. 28), which would partly buffer the $[\text{CO}_2]_{\text{aq}}$ estimates for higher cell volume and consequently for higher V/S ratios (Supplementary Information section 1).

Assuming that atmosphere and ocean are in chemical equilibrium, our $[\text{CO}_2]_{\text{aq}}$ estimates can be converted into an atmospheric p_{CO_2} of 3,200 p.p.m.v. for the upper estimate ($T = 45^\circ\text{C}$ and $\mu = 2.0\text{ d}^{-1}$) and ~ 400 p.p.m.v. for a temperature and growth rate similar to present-day values ($T = 25^\circ\text{C}$ and $\mu = 0.5\text{ d}^{-1}$) (Methods). These low p_{CO_2} estimates are in apparent contradiction to the extremely high estimates of 20,000 to 90,000 p.p.m.v. derived from boron isotope² ($\delta^{11}\text{B}$) and triple oxygen isotope³ ($\Delta^{17}\text{O}$) data. If correct, the whole data set can nonetheless be reconciled into a new environmental picture for the early Ediacaran period.

Boron isotopes, provided that the $\delta^{11}\text{B}$ of sea water is known, can be used as a proxy for seawater pH, from which atmospheric p_{CO_2} can be inferred. Kasemann *et al.*² have reported a $\sim 4\text{‰}$ decrease in $\delta^{11}\text{B}$ in the Marinoan cap carbonates of the Maieberg Formation (Namibia), which they interpret as resulting from a seawater pH decrease from 8.8 to ≤ 7 , in response to a transfer of CO_2 from a high- p_{CO_2} atmosphere into the surface ocean. As the boron isotopic composition of Neoproterozoic oceans remains unknown, however, only variations in pH can be obtained from $\delta^{11}\text{B}$ values², not the absolute seawater pH. Therefore, the reported $\delta^{11}\text{B}$ decrease is also compatible with low atmospheric p_{CO_2} , assuming a higher seawater pH. This new proposal finds additional support from the global presence of shallow-water marine carbonate in the glacial aftermath, indicating high levels of carbonate saturation, which require elevated seawater pH.

Triple oxygen isotopes measured on sulphates can also be used to constrain atmospheric p_{CO_2} (ref. 3). The acquisition of a negative $\Delta^{17}\text{O}$ signature of sulphate occurs in two steps: acquisition of a negative $\Delta^{17}\text{O}$ by the oxygen in the atmosphere, followed by transfer of the anomaly to sulphate during oxidative alteration of pyrite in exposed terrestrial sediments. The relationship between p_{CO_2} and $\Delta^{17}\text{O}(\text{O}_2)$ is given by

$$\Delta^{17}\text{O}(\text{O}_2) \propto -\Delta^{17}\text{O}(\text{CO}_{2,\text{trop}}) \frac{p_{\text{CO}_2}}{p_{\text{O}_2}} \tau_{\text{O}_2} \quad (2)$$

neglecting multiplicative constants, where $\text{CO}_{2,\text{trop}}$ is CO_2 crossing the tropopause from the stratosphere, with $\Delta^{17}\text{O}(\text{CO}_{2,\text{trop}}) \approx 1\text{‰}$ in the modern atmosphere; and τ_{O_2} is the residence time of O_2 with respect to photosynthesis and respiration (assumed to be in steady state), which is $\sim 1,200$ years in the modern atmosphere.

Recently, Bao *et al.*³ reported a negative $\Delta^{17}\text{O}$ anomaly (approximately -0.7‰) in barites intercalated in Marinoan cap carbonates, and interpreted it as a proxy for high atmospheric p_{CO_2} . Abiotic laboratory experiments and experiments with iron-oxidizing organisms have shown that 8–16% of sulphate oxygen is incorporated from atmospheric O_2 (ref. 29), with the remainder from water. Therefore, using a 10% value for atmospheric O_2 incorporation in sulphate, the observed $\Delta^{17}\text{O}(\text{barite})$ anomaly of -0.7‰ implies a $\Delta^{17}\text{O}(\text{O}_2)$ of about -7‰ . According to equation (2), $\Delta^{17}\text{O}(\text{O}_2)$ depends on two quantities: p_{CO_2} and the ratio $p_{\text{O}_2}/\tau_{\text{O}_2}$, which is the photosynthetic O_2 flux. In their model, Bao *et al.*³ assumed an atmospheric p_{O_2} of $\sim 20\%$ at the end of the Marinoan glaciation, thus resulting in a high p_{CO_2} of $\sim 10,000$ p.p.m.v. (Fig. 3, after ref. 3). We interpret these data otherwise, considering lower O_2 fluxes (that is, lower p_{O_2} and/or higher τ_{O_2}), in agreement with available evidence for low p_{O_2} levels in the early Ediacaran ($0.2\text{--}10\text{‰}$)^{30,31}. In such a case, for 1% p_{O_2} and a modern τ_{O_2} , a p_{CO_2} as low as ~ 600 p.p.m.v.—in the range of p_{CO_2} values

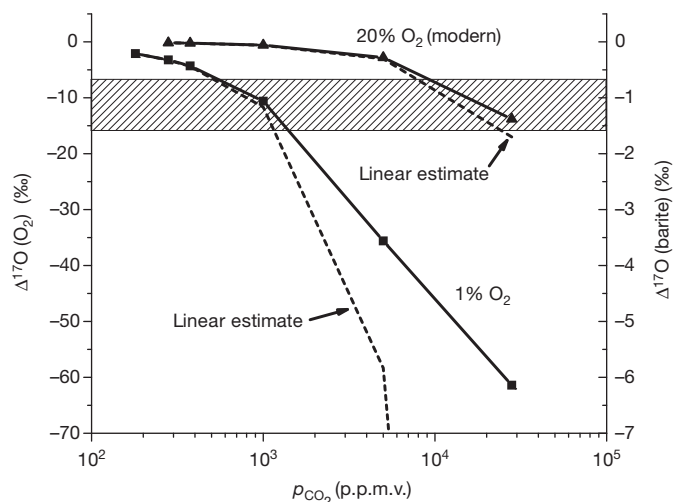


Figure 3 | Relationship between p_{CO_2} (p.p.m.v.) and $\Delta^{17}\text{O}(\text{O}_2)$ for 20% and 1% O_2 , assuming a modern value for O_2 residence time of 1,200 yr. The linear relationship between p_{CO_2} and $\Delta^{17}\text{O}(\text{O}_2)$ (dashed curves) breaks down for large $\Delta^{17}\text{O}(\text{O}_2)$ values³. The curve for 1% O_2 implies a factor of 20 reduction in the photosynthetic O_2 flux relative to the modern oceanic O_2 flux. The hatched region shows the range of maximum $\Delta^{17}\text{O}$ anomalies measured in Marinoan barites and carbonate-associated sulphates (Supplementary Information section 7). Both 1% and 20% O_2 can explain the measured $\Delta^{17}\text{O}$, but imply greatly different values for p_{CO_2} .

predicted here from paired carbon isotopic data—would be sufficient to produce the observed $\Delta^{17}\text{O}(\text{barite})$ of about -0.7‰ (Fig. 3).

In conclusion, the low $\Delta^{13}\text{C}_{\text{carb-orig}}$ values reported here for Marinoan cap carbonates not only provide a lower p_{CO_2} estimate than previously thought, but also allow the re-interpretation of $\delta^{11}\text{B}$ and $\Delta^{17}\text{O}$ isotopic data. The new environmental picture deduced from this integrated interpretation, with low atmospheric p_{CO_2} , high seawater pH and low atmospheric p_{O_2} (associated with low O_2 fluxes), represents a substantial challenge to the hard-end-member snowball Earth picture¹.

METHODS SUMMARY

Samples were ground in an agate mortar, then sieved to ensure a grain size of $< 140\text{ }\mu\text{m}$. For carbonate isotope analyses, we used 100% H_3PO_4 to extract CO_2 successively from calcite and dolomite, in a two-step dissolution process (4 h at 25°C for calcite, then 2 h at 80°C for dolomite). We measured carbon and oxygen isotope compositions of the evolved CO_2 using a gas chromatograph coupled to a GV Instruments Analytical Precision 2003 mass spectrometer. The external reproducibilities (1σ) for $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{18}\text{O}_{\text{carb}}$ measurements are 0.1‰ and 0.2‰ , respectively. For C and N quantification, and organic carbon analysis, samples were decarbonated in 6 N HCl overnight at room temperature, followed by 2 h at 80°C . Residues were washed with distilled water, centrifuged and dried at 50°C . Samples of decarbonated powder (10–60 mg) were loaded into quartz tubes along with copper oxide wires. The tubes were connected to a vacuum line and sealed under secondary vacuum ($< 10^{-5}$ mbar), then heated at 950°C for 6 h. The resulting CO_2 and N_2 were purified on a vacuum line and manometrically quantified using a Toepler pump. Total organic carbon content, nitrogen content and hence C/N are deduced from the CO_2 and N_2 quantification with a precision of $\pm 10\%$ relative to the measured value. The carbon isotope composition was measured using a dual-inlet Thermo Finnigan Delta+XP mass spectrometer. The reproducibility of the $\delta^{13}\text{C}_{\text{org}}$ measurement is $\pm 0.1\text{‰}$ (1σ). All isotopic results are given in δ notation calibrated to V-PDB (Vienna Pee Dee Belemnite). Rock-Eval analyses and organic geochemistry were performed using standard methods, described in the online Methods section.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 May 2010; accepted 22 August 2011.

- Hoffman, P. F. & Schrag, D. P. The snowball Earth hypothesis: testing the limits of global change. *Terra Nova* **14**, 129–155 (2002).

2. Kasemann, S. A., Hawkesworth, J. C., Prave, A. R., Fallick, A. E. & Pearson, P. N. Boron and calcium isotope composition in Neoproterozoic carbonate rocks from Namibia: evidence for extreme environmental change. *Earth Planet. Sci. Lett.* **231**, 73–86 (2005).
3. Bao, H., Lyons, J. R. & Zhou, C. Triple oxygen isotope evidence for elevated CO₂ levels after a Neoproterozoic glaciation. *Nature* **453**, 504–506 (2008).
4. Shen, B. *et al.* Stratification and mixing of a post-glacial Neoproterozoic ocean: Evidence from carbon and sulfur isotopes in a cap dolostone from northwest China. *Earth Planet. Sci. Lett.* **265**, 209–228 (2008).
5. McFadden, K. A. *et al.* Pulsed oxidation and biological evolution in the Ediacaran Doushantuo Formation. *Earth Planet. Sci. Lett.* **105**, 3197–3202 (2007).
6. Jiang, G. *et al.* Organic carbon isotope constraints on the dissolved organic carbon (DOC) reservoir at the Cryogenian-Ediacaran transition. *Earth Planet. Sci. Lett.* **299**, 159–168 (2010).
7. Guo, Q. *et al.* Carbon isotopic evolution of the terminal Neoproterozoic and early Cambrian: Evidence from the Yangtze Platform, South China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **254**, 140–157 (2007).
8. Ader, M. *et al.* A multilayered water column in the Ediacaran Yangtze platform? Insights from carbonate and organic matter paired $\delta^{13}\text{C}$. *Earth Planet. Sci. Lett.* **288**, 213–227 (2009).
9. Hayes, J. M., Strauss, H. & Kaufman, A. J. The abundance of ^{13}C in marine organic matter and isotopic fractionation in the global biogeochemical cycle of carbon during the past 800 Ma. *Chem. Geol.* **161**, 103–125 (1999).
10. Kaufman, A. J. & Xiao, S. High CO₂ levels in the Proterozoic atmosphere estimated from analyses of individual microfossils. *Nature* **425**, 279–282 (2003).
11. Young, S. A., Saltzman, M. R., Bergström, S. M., Leslie, S. A. & Xu, C. Paired $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{13}\text{C}_{\text{org}}$ records of Upper Ordovician (Sandbian-Katian) carbonates in North America and China: Implications for paleoceanographic change. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **270**, 166–178 (2008).
12. Riccardi, A., Kump, L. R., Arthur, M. A. & D'Hondt, S. Carbon isotopic evidence for chemocline upward excursions during the end-Permian event. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **248**, 73–81 (2007).
13. de Alvarenga, C. J. S., Santos, R. V. & Dantas, E. L. C-O-Sr isotopic stratigraphy of cap carbonates overlying Marinoan-age glacial diamictites in the Paraguay Belt, Brazil. *Precamb. Res.* **131**, 1–21 (2004).
14. Corsetti, F. A. & Kaufman, A. J. Stratigraphic investigations of carbon isotope anomalies and Neoproterozoic ice ages in Death Valley, California. *Geochim. Cosmochim. Acta* **115**, 916–932 (2003).
15. Kaufman, A. J., Hayes, J. M., Knoll, A. H. & Germs, G. J. B. Isotopic compositions of carbonates and organic carbon from upper Proterozoic successions in Namibia: stratigraphic variation and the effect of diagenesis and metamorphism. *Precamb. Res.* **49**, 301–327 (1991).
16. Narbonne, G. M., Kaufman, A. J. & Knoll, A. H. Integrated chemostratigraphy and biostratigraphy of the Windermere Supergroup, northwestern Canada: Implications for Neoproterozoic correlations and the early evolution of animals. *Geol. Soc. Am. Bull.* **106**, 1281–1292 (1994).
17. Knauth, L. P. & Kennedy, M. J. The late Precambrian greening of the Earth. *Nature* **460**, 728–732 (2009).
18. Bristow, T. F. & Kennedy, M. Carbon isotope excursions and the oxidant budget of the Ediacaran atmosphere and ocean. *Geol. Soc. Am.* **36**, 863–866 (2008).
19. Font, E., Trindade, R. I. F. & Nédélec, A. Detrital remanent magnetization in haematite-bearing Neoproterozoic Puga cap dolostone, Amazon craton: a rock magnetic and SEM study. *Geophys. J. Int.* **163**, 491–500 (2005).
20. Crafts-Brandner, S. J. & Salvucci, M. E. Rubisco activase constrains the photosynthetic potential of leaves at high temperature and CO₂. *Proc. Natl Acad. Sci. USA* **97**, 13430–13435 (2000).
21. Huber, R., Huber, H. & Stetter, K. O. Towards the ecology of hyperthermophiles: biotopes, new isolation strategies and novel metabolic properties. *FEMS Microbiol. Rev.* **24**, 615–623 (2000).
22. Corsetti, F. A., Olcott, A. N. & Bakermans, C. The biotic response to Neoproterozoic snowball Earth. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **232**, 114–130 (2006).
23. Elie, M., Nogueira, A. C. R., Nédélec, A., Trindade, R. I. F. & Kenig, F. A red algal bloom in the aftermath of the Marinoan snowball Earth. *Terra Nova* **19**, 303–308 (2007).
24. Goericke, R. & Fry, B. Variations of marine plankton $\delta^{13}\text{C}$ with latitude, temperature, and dissolved CO₂ in the world ocean. *Glob. Biogeochem. Cycles* **8**, 85–90 (1994).
25. Rost, B., Riebesell, U., Burkhardt, S. & Sültermeier, D. Carbon acquisition of bloom-forming marine phytoplankton. *Limnol. Oceanogr.* **48**, 55–67 (2003).
26. Werne, J. P. & Hollander, D. J. Balancing supply and demand: controls on carbon isotope fractionation in the Cariaco Basin (Venezuela) Younger Dryas to present. *Mar. Chem.* **92**, 275–293 (2004).
27. Popp, B. N. *et al.* Effect of phytoplankton cell geometry on carbon isotopic fractionation. *Geochim. Cosmochim. Acta* **62**, 69–77 (1998).
28. Finkel, Z. V. *et al.* Phytoplankton in a changing world: cell size and elemental stoichiometry. *J. Plankton Res.* **32**, 119–137 (2010).
29. Balci, N., Shanks, W. C., Mayer, B. & Mandernack, K. W. Oxygen and sulfur isotope systematics of sulfate produced by bacterial and abiotic oxidation of pyrite. *Geochim. Cosmochim. Acta* **71**, 3796–3811 (2007).
30. Rye, R. & Holland, H. D. Paleosols and the evolution of atmospheric oxygen: A critical review. *Am. J. Sci.* **298**, 621–672 (1998).
31. Li, C. *et al.* A stratified redox model for the Ediacaran ocean. *Science* **328**, 80–83 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Geochemistry division of IFP Energie nouvelle for performing Rock-Eval analyses. The work benefited from discussions with M. Bonifacie, G. LeHir, G. Paris and H. Strauss. Research was supported by a French MRT doctoral fellowship and a SETSI grant to P.S., and two INSU (SYSTER) grants to M.A., R.I.F.T. and A.C.R.N. were supported by the INCT-Geociam programme, and by FAPESP and CNPq grants. J.L. was supported by the NASA Astrobiology Institute under cooperative agreement NNA09DA76 to the Penn State Astrobiology Research Center. This is IGPp contribution 3211.

Author Contributions M.A. and R.I.F.T. conceived the work. P.S., M.A., R.I.F.T. and A.C.R.N. did the sampling. P.S., M.A., R.I.F.T. and P.C. wrote the paper and most of the Supplementary Information. P.S. carried out carbon isotope analyses. P.S. and M.E. did molecular organic geochemistry analyses and wrote the related Supplementary Information. M.A. and M.E. wrote Supplementary Information corresponding to the Rock-Eval data. J.L. performed triple-oxygen modelling and wrote the corresponding parts of the main text and Supplementary Information. A.C.R.N. organized the field work and contributed the geological setting of samples. All authors discussed results and contributed to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.S. (sansjofre@ipgp.fr).

METHODS

p_{CO_2} reconstruction. As explained in the main text, ε_p is calculated from $\Delta^{13}\text{C}_{\text{carb-org}}$ knowing Δ_{carb} and Δ_2 (fixed at 1.5; Supplementary Information section 5.1). Δ_{carb} can be decomposed as the sum of the isotope fractionation factor between carbonate and HCO_3^- ($\Delta^{13}\text{C}_{\text{carb-HCO}_3^-}$) and the one between HCO_3^- and $\text{CO}_{2\text{aq}}$ ($\Delta^{13}\text{C}_{\text{HCO}_3^--\text{CO}_{2\text{aq}}}$). Three sources of $\Delta^{13}\text{C}_{\text{carb-HCO}_3^-}$ are available: the $\Delta^{13}\text{C}_{\text{CaCO}_3-\text{HCO}_3^-}$ for inorganic calcite precipitation of 0.9‰ at ambient temperature³²; the $\Delta^{13}\text{C}_{\text{dolomite-HCO}_3^-}$, which ranges from 3.3‰ to 1.2‰ at ambient temperature³³; and the $\Delta^{13}\text{C}_{\text{calcite-HCO}_3^-}$ of -1.2 ‰, estimated in ref. 9 from the difference between modern carbonate sediments and surface seawater HCO_3^- . We chose this smallest value of $\Delta^{13}\text{C}_{\text{CaCO}_3-\text{HCO}_3^-}$, to ensure that the ultimately derived p_{CO_2} estimate represents an upper limit. Since $\Delta^{13}\text{C}_{\text{CaCO}_3-\text{HCO}_3^-}$ may change slightly with temperature, we used the variation of 0.01 ‰ °C⁻¹ also proposed in ref. 9. In contrast, $\Delta^{13}\text{C}_{\text{HCO}_3^--\text{CO}_{2\text{aq}}}$ is strongly temperature-dependent and is calculated using the formula $\Delta^{13}\text{C}_{\text{HCO}_3^--\text{CO}_{2\text{aq}}} = 9866/(T+273) - 24.12$ (ref. 34).

$[\text{CO}_2]_{\text{aq}}$ is calculated from ε_p using the relation obtained by Popp *et al.*²⁷ from cultures of unicellular algae (Fig. 2). Other $[\text{CO}_2]_{\text{aq}}-\varepsilon_p$ equations have been determined but most of them were obtained for single species at a local scale (in lakes) and do not take into account variations in physiological parameters³⁵. Equations based on mixed phytoplankton populations^{36,37}, which take into account V/S and μ , yield $[\text{CO}_2]_{\text{aq}}$ estimates inferior to those obtained using Popp *et al.*'s equation. Hence, the choice of this equation provides an upper estimate of $[\text{CO}_2]_{\text{aq}}$. Moreover, the equation used here is more compatible with the algal signature of molecular organic data (Supplementary Information section 4).

Finally, $[\text{CO}_2]_{\text{aq}}$ is converted into p_{CO_2} , assuming that ocean and atmosphere were at equilibrium as they are today²⁴. Although episodic events of high growth rate can drive the ocean-atmosphere system out of equilibrium owing to the rapid drawdown of ocean CO_2 by photosynthesizers, they are typically local and occur on short timescales. The ubiquitous and homogeneous isotopic record of cap carbonates points instead to a global and persistent process. We used Henry's law: $p_{\text{CO}_2} = [\text{CO}_2]_{\text{aq}}/k_0$. The Henry constant (k_0) is expressed as a function of temperature (T) and salinity (S)³⁸ as follows: $\ln(k_0) = 9345.17/T - 60.2409 + 23.3585 \ln(T/100) + S[0.023517 - 0.00023656T + 0.0047036(T/100)^2]$.

We have tested salinity values from 15 to 50 μM (modern value being 35 μM). In the main text, we present only the value corresponding to the highest p_{CO_2} estimate obtained with a high salinity of 50 μM .

Carbon isotope analysis, C and N quantification. Samples were ground in an agate mortar, then sieved to ensure a grain size of <140 μm . Powdered samples were reacted with 100% H_3PO_4 at 25 °C for 4 h to extract the CO_2 from calcite, and then at 80 °C for 2 h to extract CO_2 from dolomite. Carbon and oxygen isotope compositions were measured using a helium continuous flow mass spectrometer (AP 2003). Isotopic compositions are given in the δ notation relative to the V-PDB (Vienna Pee Dee Belemnite). The external reproducibilities (1σ) for $\delta^{13}\text{C}_{\text{carb}}$ and $\delta^{18}\text{O}_{\text{carb}}$ measurements are 0.1‰ and 0.2‰, respectively. $\delta^{13}\text{C}_{\text{carb}}$ values measured on calcite and dolomite were similar, except in the lower part of the dolomite where calcite $\delta^{13}\text{C}_{\text{carb}}$ values were lower by a maximum of 2‰. Because petrographic observations of these dolostones attest to the secondary character of calcite (Supplementary Information section 3), only the dolomite $\delta^{13}\text{C}_{\text{carb}}$ values were considered. For C and N quantification, and organic carbon analysis,

samples were decarbonated in 6N HCl overnight at room temperature, followed by 2 h at 80 °C. Residues were washed with distilled water, centrifuged and dried at 50 °C. Samples of decarbonated powder (10–60 mg) were loaded into quartz tubes along with copper oxide wires. The tubes were connected to a vacuum line and sealed under secondary vacuum ($<10^{-5}$ mbar), then heated at 950 °C for 6 h. The resulting CO_2 and N_2 were purified on a vacuum line and manometrically quantified using a Toepler pump. Total organic carbon content, nitrogen content and hence C/N are deduced from the CO_2 and N_2 quantification with a precision of $\pm 10\%$ relative to the measured value. The carbon isotope composition was measured using a dual-inlet Thermo Finnigan Delta+XP mass spectrometer at the IPGP, and is expressed in δ notation calibrated to V-PDB (Vienna Pee Dee Belemnite). The reproducibility of the $\delta^{13}\text{C}_{\text{org}}$ measurement is ± 0.1 ‰ (1σ).

Molecular organic geochemistry. Analyses were performed at Henri-Poincaré University of Nancy. The soluble organic matter was extracted with dichloromethane at 100 bar and 80 °C using an Accelerated Solvent Extractor ASE 200. A blank was performed before each extraction. Two extraction cycles were performed to ensure a complete extraction. Elemental S was removed by introducing HCl activated Cu chips in vials containing the solvent and the extract. Dichloromethane was evaporated using a Zymark TurboVap LV. The extracted organic matter was fractionated into aliphatic and aromatic hydrocarbons on a silica column by successive elution of pentane and pentane/dichloromethane (65/35). Aliphatic hydrocarbons were diluted in hexane (4 mg ml⁻¹) and analysed on a HP5890 Serie II Gas chromatograph coupled with a HP5971 Mass Spectrometer (GC-MS) following the procedure described in ref. 23.

Rock-Eval analyses. Because a TOC content of 0.3% is required for a meaningful determination of Rock-Eval parameters, the bulk-rock organic matter was first concentrated by HF and HCl mineral dissolution using a kerogenatron. Analyses were carried out on the organic concentrate using a Rock-Eval 6 Turbo device at the Institut Français du Pétrole following the classical methodology³⁹. The Rock-Eval parameters used were the hydrogen index (HI, mg HC per g TOC) and oxygen index (OI, mg CO_2 per g TOC) also described in ref. 39.

32. Zeebe, R. E. & Wolf-Gladrow, D. *CO₂ in Seawater: Equilibrium, Kinetics, Isotopes* (Elsevier Oceanography Series **65**, (2001).
33. Sheppard, S. & Schwarz, H. Fractionation of carbon and oxygen isotopes and magnesium between coexisting metamorphic calcite and dolomite. *Contrib. Mineral. Petrol.* **26**, 161–198 (1970).
34. Mook, W. G., Bommerson, C. J. & Staberman, W. H. Carbon isotope fractionation between dissolved bicarbonate and gaseous carbon dioxide. *Earth Planet. Sci. Lett.* **22**, 169–176 (1974).
35. Royer, D. L., Berner, R. A. & Beerling, D. J. Phanerozoic atmospheric CO_2 change: evaluating geochemical and paleobiological approaches. *Earth Sci. Rev.* **54**, 349–392 (2001).
36. Laws, E. A., Popp, B. N., Bidigare, R. R., Kennicutt, M. C. & Macko, S. A. Dependence of phytoplankton carbon isotopic composition on growth rate and $[\text{CO}_2]_{\text{aq}}$: Theoretical considerations and experimental results. *Geochim. Cosmochim. Acta* **59**, 1131–1138 (1995).
37. Pancost, R. D., Freeman, K. H. & Wakeham, S. G. Controls on the carbon-isotope compositions of compounds in Peru surface waters. *Org. Geochem.* **30**, 319–340 (1999).
38. Weiss, R. F. Carbon dioxide in water and seawater: the solubility of a non-ideal gas. *Mar. Chem.* **2**, 203–215 (1974).
39. Béhar, F., Beaumont, V., De, H. L. & Pentead, B. Rock-Eval 6 technology: performances and developments. *Rev. Inst. Fr. Pet.* **56**, 111–134 (2001).

Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus

A list of authors and their affiliations appears at the end of the paper

Both obesity and being underweight have been associated with increased mortality^{1,2}. Underweight, defined as a body mass index (BMI) ≤ 18.5 kg per m² in adults and ≤ -2 standard deviations from the mean in children, is the main sign of a series of heterogeneous clinical conditions including failure to thrive^{3–5}, feeding and eating disorder and/or anorexia nervosa^{6,7}. In contrast to obesity, few genetic variants underlying these clinical conditions have been reported^{8,9}. We previously showed that hemizygosities of a ~600-kilobase (kb) region on the short arm of chromosome 16 causes a highly penetrant form of obesity that is often associated with hyperphagia and intellectual disabilities¹⁰. Here we show that the corresponding reciprocal duplication is associated with being underweight. We identified 138 duplication carriers (including 132 novel cases and 108 unrelated carriers) from individuals clinically referred for developmental or intellectual disabilities (DD/ID) or psychiatric disorders, or recruited from population-based cohorts. These carriers show significantly reduced postnatal weight and BMI. Half of the boys younger than five years are underweight with a probable diagnosis of failure to thrive, whereas adult duplication carriers have an 8.3-fold increased risk of being clinically underweight. We observe a trend towards increased severity in males, as well as a depletion of male carriers among non-medically ascertained cases. These features are associated with an unusually high frequency of selective and restrictive eating behaviours and a significant reduction in head circumference. Each of the observed phenotypes is the converse of one reported in carriers of deletions at this locus. The phenotypes correlate with changes in transcript levels for genes mapping within the duplication but not in flanking regions. The reciprocal impact of these 16p11.2 copy-number variants indicates that severe obesity and being underweight could have mirror aetiologies, possibly through contrasting effects on energy balance.

Copy-number variants (CNVs) at the 16p11.2 locus have been associated with cognitive disorders including autism (deletions) and schizophrenia (duplications)^{11–13}, conditions that have been suggested to lie at opposite ends of a single spectrum of psychiatric phenotypes¹⁴. We and others have reported that a deletion of this region spanning 28 genes (Supplementary Table 1) increases the risk of morbid obesity 43-fold (Supplementary Fig. 1)^{10,15}. We hypothesized that the reciprocal duplication, with its resulting increase in gene dosage, may influence BMI in a converse manner. The duplication was identified in 73 out of 31,424 patients with DD/ID, a frequency consistent with previous reports¹³ (Table 1). Four additional cases were identified among 1,080 patients affected by bipolar disease or schizophrenia. Compared to its prevalence in seven European population-based genome-wide association study (GWAS) cohorts^{16–18} (31 out of 58,635 individuals), the duplication was significantly more frequent in both the DD/ID cohorts ($P = 4.23 \times 10^{-13}$; odds ratio = 4.4, 95% confidence interval = 2.9–6.9) and the psychiatric cohorts ($P = 3.6 \times 10^{-3}$; odds ratio = 7.0, 95% confidence interval = 1.8–19.9) (Table 1), strengthening previous reports of similar associations^{12,13}. Our data do not support a two-hit model¹⁹ for the effects of 16p11.2 duplications or deletions (Supplementary Text and Supplementary Table 2).

We compared available data on weight, height and BMI for 106 independent duplication carriers (including published cases) to data for reference populations matched for gender, age and geographical location (Table 2, Methods and Supplementary Tables 3 and 4). The duplication was strongly associated with lower weight (mean Z-score -0.56 ; $P = 4.4 \times 10^{-4}$) and lower BMI (mean Z-score -0.47 ; $P = 2.0 \times 10^{-3}$) (Table 2 and Supplementary Table 5). Birth parameters ($n = 48$) were normal, indicating a postnatal effect. Adults carrying the duplication had a relative risk of being clinically underweight (BMI < 18.5) of 8.3 (95% confidence interval = 4.4–15.9, $P = 1.53 \times 10^{-10}$) (see Methods). Concordantly, none of the 3,544 patients in our obesity cohorts^{10,15} carried the duplication (Table 1).

To investigate these associations further, we carried out separate analyses of carrier patients (DD/ID and psychiatric) and non-medically ascertained carriers (population-based cohorts, plus 11 transmitting parents and three other affected first-degree relatives for whom data were available) (Table 2). Each category had significantly lower weight and BMI, with similar effect sizes. However, the proportion of underweight cases (BMI ≤ -2 s.d.) was higher in the first group than in the second group (17 out of 76 compared to 2 out of 40; $P = 0.017$). Note that the impact of the duplication on underweight status might be underestimated here owing to prescription of antipsychotic treatments that are often associated with weight gain²⁰ (Supplementary Table 6).

Having demonstrated an association of the duplication with being underweight, we investigated the implications of gender for the resulting phenotypes (Fig. 1, Supplementary Fig. 2 and Supplementary Table 7). In DD/ID patients, the impact of the duplication on being underweight is stronger in males; the effect in females is in the same direction, but is smaller and not statistically significant (Table 2). A similar and significant difference ($P = 0.0168$) was observed in adult carriers (all groups combined): the relative risk of being underweight for males is 23.2 (95% confidence interval = 9.1–59.3, $P = 4.6 \times 10^{-11}$); for females it is only 4.7 (95% confidence interval = 1.9–11.8, $P = 9.9 \times 10^{-4}$). A gender bias was also observed in the ascertainment of DD/ID duplication carriers, in which we have an excess of males (51 males:33 females, $P = 0.044$). By contrast, carriers from the general population showed a strong overrepresentation of females (10 males:21 females, $P = 0.035$) (Supplementary Text). A similar bias was observed among transmitting parents (7 males:23 females, $P = 5.53 \times 10^{-4}$). Thus, there is an overrepresentation of males in the medically ascertained group, and a depletion in the non-medically ascertained one. We suggest that males may be more likely than females to present severe phenotypes, and that this may account for the gender bias because severely affected males may be less likely to be recruited to adult population cohorts or to be reproductively successful.

As previously reported²¹, the duplication was also associated with reduced head circumference (mean Z-score -0.89 ; $P = 7.8 \times 10^{-6}$) (Fig. 1), 26.7% presenting with microcephaly (head circumference ≤ -2 s.d.), whereas carriers of the reciprocal deletion had an increased head circumference (mean Z-score $+0.57$; $P = 1.79 \times 10^{-5}$) (Supplementary Fig. 3 and Supplementary Table 8): an additional instance of a mirror phenotype associated with reciprocal changes in copy number at this locus. Notably, head circumference Z-scores correlate

Table 1 | 16p11.2 rearrangements in cases and controls

Ascertainment	Cohorts	Duplication		Deletion		Total
		<i>n</i>	<i>P</i> †	<i>n</i>	<i>P</i> †	
Neuro-developmental disorders	Unspecified DD/ID* from 28 cytogenetic centres	72		113		30,323
	ADHD‡, deCODE	0		1		591
	Childhood autism‡, deCODE	0		2		159
	Childhood autism spectrum disorder‡, deCODE	1		3		351
	TOTAL	73	4.23 × 10⁻¹³	119	5.43 × 10⁻³²	31,424
Family history	Rearrangement frequency (95% CI)	0.23% (0.18–0.29)		0.38% (0.31–0.45)		
	First-degree relatives of probands	30		35		43/62
	Schizophrenia, deCODE	0		1		657
	Bipolar disease, Rouen	1		0		156
	Schizophrenia, schizo-affective, Rouen	3		0		267
Adult psychiatric symptoms	TOTAL	4	3.57 × 10⁻³	1	3.78 × 10⁻¹	1,080
	Rearrangement frequency (95% CI)	0.37% (0.01–0.73)		0.09% (0–0.27)		
	Eating disorder, Spain	1§		0		441
	Obesity, Spain	0		2		653
	Adult obesity, France	0		4		705
Underweight	Childhood obesity, France & UK	0		7		1,574
	Obesity bariatric surgery, France	0		2		141
	Obesity discordant siblings, Sweden	0		2		159
	Obesity and cognitive delay, France & UK	0		9		312
	TOTAL	0	4.21 × 10⁻¹	26	2.52 × 10⁻¹⁹	3,544
Obesity	Rearrangement frequency (95% CI)	0		0.73% (0.45–1.01)		
	NFBC1966 Finnish	4		3		5,319
	CoLaus Swiss	5		0		5,612
	EGCUT Estonian	2		1		2,994
	deCODE Iceland	17		18		36,601
Population-based cohorts	SHIP Germany	1		2		4,070
	KORA F3 + F4 Germany	2		1		3,458
	Paediatric family study	0		0		581
	TOTAL	31		25		58,635
	Rearrangement frequency (95% CI)	0.05% (0.03–0.07)		0.04% (0.03–0.06)		

CI, Confidence interval; ADHD, attention-deficit hyperactivity disorder. *Not a disease-specific cohort. Detailed distribution is provided in the online methods. †Fisher's exact test, compared to the combined frequency in general population groups. ‡There was no overlap between these 3 cohorts. §Atypical duplication (see Supplementary Fig. 5). || Total number of parental pairs tested for duplication/deletion. 13 out of 43 duplications and 27 out of 62 deletion cases were *de novo*.

positively with those of BMI in carriers of both the duplication ($\rho = 0.37$; $P = 2.65 \times 10^{-3}$) and the deletion ($\rho = 0.42$; $P = 1.9 \times 10^{-5}$) (Supplementary Methods). This indicates that head circumference and BMI may be regulated by a common pathway, or that a causal relationship exists between these two traits in these patients. Alternatively, the two phenotypes may arise from distinct genes and pathways. A full list of malformations and secondary phenotypes reported in duplication carriers ascertained for DD/ID is available in Supplementary Table 9.

In view of the importance of modified eating behaviours in obesity and being underweight, the clinical reports of duplication carriers were screened for evidence of such modified behaviours. In 11 out of 77 clinically ascertained cases, clinicians had spontaneously reported low food intake and selective and restrictive eating behaviour, again mirroring one of the phenotypes—hyperphagia—seen in deletion carriers¹⁰ (Supplementary Table 6) and indicating that the duplication may increase the risk of eating disorders. Consequently, we carried out

multiplex ligation-dependent probe amplification (MLPA, Supplementary Table 10) to screen for 16p11.2 rearrangements in 441 patients diagnosed with eating disorders, including anorexia nervosa, bulimia and binge eating disorder (Table 1 and Supplementary Text). No duplications of the entire region were identified, but one out of 109 anorexia nervosa patients carried an atypical 136-kb duplication that encompasses the sialophorin (*SPN*) and quinolinate phosphoribosyltransferase (*QPRT*) genes (Supplementary Fig. 4). This single, smaller duplication does not allow us to draw any firm conclusions, but together with other atypical rearrangements, it may, in the future, be essential for establishing the roles of the 28 genes within the region.

Large genomic structural variants are known to affect the expression of genes not only within the affected region but also at a distance^{22–25}. Therefore, it is possible that the phenotypes observed in 16p11.2 deletion and duplication individuals are due to effects on the expression of genes mapping outside the rearranged interval, rather than to gene dosage within the 600-kb deletion or duplication. We measured

Table 2 | Comparisons of the height, weight and BMI distributions in duplication carriers and controls.

	Strata	Combined†			DD/ID or psychiatric‡			Non-medically ascertained‡		
		Mean Z-score	<i>P</i> -value	<i>n</i> *	Mean Z-score	<i>P</i> -value	<i>n</i> *	Mean Z-score	<i>P</i> -value	<i>n</i> *
BMI	All	−0.47	2.0 × 10⁻³	102	−0.56	4.1 × 10⁻³	76	−0.45	6.0 × 10⁻³	40
	Male	−0.54	2.1 × 10⁻²	52	−0.71	1.3 × 10⁻²	43	−0.31	2.0 × 10⁻¹	14
	Female	−0.4	1.8 × 10⁻²	50	−0.37	8.3 × 10⁻²	33	−0.52	4.2 × 10⁻³	26
Weight	All	−0.56	4.4 × 10⁻⁴	104	−0.65	1.3 × 10⁻³	78	−0.61	3.0 × 10⁻³	40
	Male	−0.64	5.8 × 10⁻³	53	−0.79	4.4 × 10⁻³	44	−0.57	8.8 × 10⁻²	14
	Female	−0.47	1.7 × 10⁻²	51	−0.47	6.5 × 10⁻²	34	−0.63	8.6 × 10⁻³	26
Height	All	−0.24	4.8 × 10⁻²	103	−0.33	3.6 × 10⁻²	77	−0.15	1.8 × 10⁻¹	40
	Male	−0.34	4.5 × 10⁻²	52	−0.4	4.6 × 10⁻²	43	−0.29	1.2 × 10⁻¹	14
	Female	−0.14	2.6 × 10⁻¹	51	−0.24	2.1 × 10⁻¹	34	−0.07	3.7 × 10⁻¹	26

The available BMI, weight and height data for duplication carriers were transformed to Z-scores using gender- and age-matched reference populations, and one-tailed *t*-tests were carried out to determine whether the mean Z-scores deviated from zero. Significant differences were identified by reference to cutoffs controlling the false discovery rate at 5% (see Methods): BMI, 0.022; weight, 0.032; height, 0.025. Significant results are indicated in bold. Data were not available for all subjects. *Relatives of probands were excluded as required, to avoid including more than one member of the same family in a single analysis. †Including 24 cases from the literature (Supplementary Table 3). ‡Population-based cases and first-degree relatives of probands.

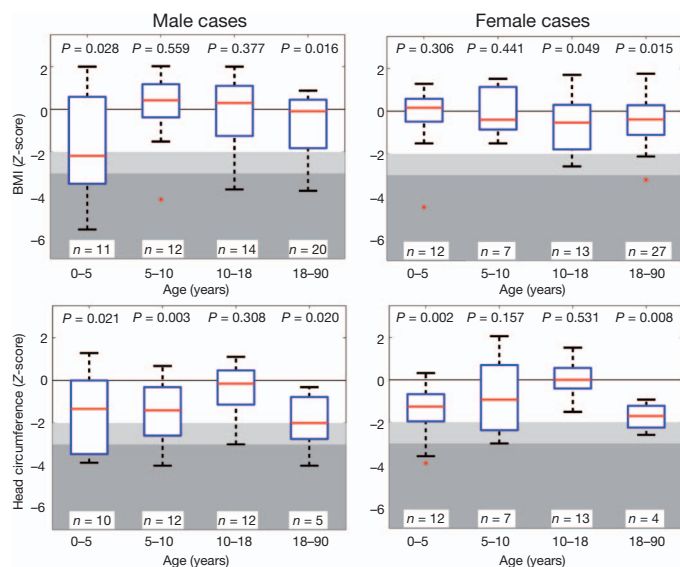


Figure 1 | Effect of the chromosome 16p11.2 duplication on BMI and head circumference. Z-score values of BMI and head circumference in carriers of the 16p11.2 duplication, stratified by gender and age group. The most severe effect is observed in children at 0–5 years of age. Boxplots represent the fifth, twenty-fifth, median, seventy-fifth and ninety-fifth percentile for each age group. Light grey and dark grey backgrounds represent ≤ -2 and ≤ -3 s.d., respectively, corresponding to the WHO definition of moderately and severely underweight³⁰. BMI is decreased in adolescent and adult females.

relative transcript levels of 27 genes mapping within or near to the rearrangement, using lymphoblastoid cell lines (Supplementary Tables 1 and 11): six from deletion carriers, five from duplication carriers and ten from gender- and age-matched controls (Supplementary Table 12). Expression levels correlated positively with gene dosage for all genes in the CNV region (Fig. 2), consistent with published partial results from adipose tissue¹⁰. Mean relative transcript levels in deletion and duplication carriers were, respectively, 67% and 214% of the levels measured in controls (Supplementary Table 13). Although genes proximal (centromeric) to the rearrangement interval

showed no significant variation in relative transcript levels between patients and controls (Fig. 2), distal (telomeric) genes showed a marked alteration in relative expression. However, their expression levels, including that of *SH2B1* (for which gene dosage and a nearby single nucleotide polymorphism (SNP) have been associated with obesity^{15,26}), were similarly upregulated in cell lines of both deletion and duplication carriers, showing no apparent correlation between transcript level and either copy number or phenotype (Fig. 2). Although lymphoblastoid cells may not recapitulate obesity-relevant tissues, previous experiments have shown a high degree of correlation between expression levels in different tissues and cell lines²², indicating that the same pathways may be similarly disrupted in different cell lineages. Thus, any involvement of these distal genes in the control of BMI in these subjects seems unlikely.

Our study demonstrates the power of very large screens ($>95,000$ samples: to our knowledge the largest of its kind so far) to characterize the clinical and molecular correlates of a rare functional genomic variant. We demonstrate unambiguously that carrying the 16p11.2 duplication confers a high risk of being clinically underweight, and show that reciprocal changes in gene dosage at this locus result in several mirror phenotypes. As in the schizophrenia/autism¹⁴ and microcephaly/macrocephaly²¹ dualisms, abnormal eating behaviours, such as hyperphagia and anorexia, could represent opposite pathological manifestations of a common energy-balance mechanism, although the precise relationships between these mirror phenotypes remain to be determined. We speculate that head circumference (which correlates with brain volume²⁷), and thus neuronal circuitry, may affect cognitive function and energy balance in patients with 16p11.2 rearrangements (possibly through eating behaviour). Consistent with this are previous reports that a subgroup of children with microcephaly show a concomitant reduction in weight percentile²⁸. Our findings also support the observation that severe overweight and underweight phenotypes correlate with lower cognitive functioning^{4,29}. Thus, abnormal food intake may be a direct result of particular neurodevelopmental disorders. Although it is possible that the 16p11.2 region encodes distinct genes specific for each trait, a more parsimonious hypothesis is that these clinical manifestations of dysfunction of the central nervous system are all secondary to the disruption of a single neurodevelopmental step that is sensitive to gene dosage. Further resolution of this issue may

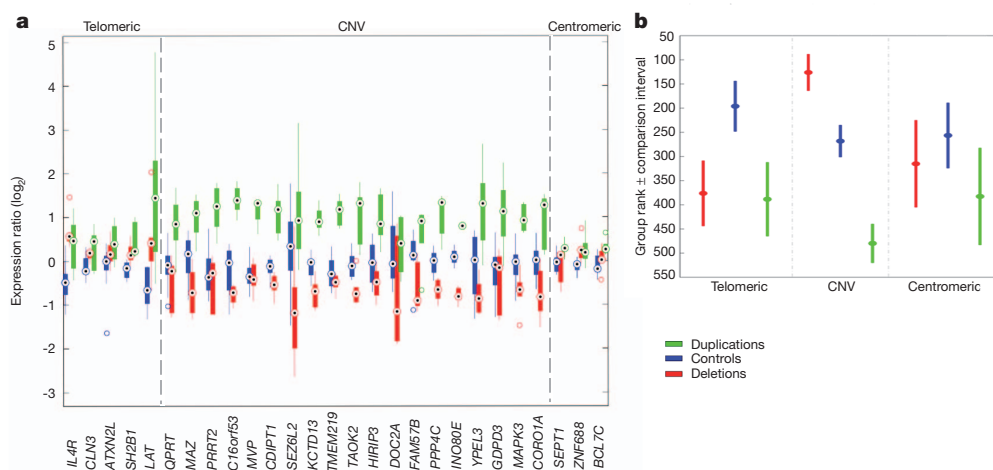


Figure 2 | Transcript levels for genes within and near to the 16p11.2 rearrangements. **a**, Relative expression levels of 27 genes mapping to 16p11.2 in deletion ($n = 6$) and duplication ($n = 5$) carriers (red and green, respectively), and in control cell lines ($n = 10$, blue). Grey lines denote the extent of the 16p11.2 CNV (29.5–30.1 megabases (Mb)). Complete lists of genes mapping within the rearranged interval, and of the quantitative PCR assays, are in Supplementary Tables 1 and 11, respectively. For the possible relevance of each of these genes to obesity/leanness and/or developmental delay/cognitive deficits, see ref. 10. **b**, Rank comparison (Kruskal–Wallis test)

between the expression of 27 genes mapping to 16p11.2 in deletion and duplication carriers (red and green, respectively) and in control cell lines (blue). Genes are labelled as telomeric, centromeric or within the rearranged interval (CNV). Dots correspond to the mean group rank and bars indicate the comparison interval. Groups with non-overlapping intervals are significantly different (P -values were adjusted for multiple testing issues using a Bonferroni correction, where the number of tests is the number of pairwise comparisons; the resulting adjusted P -value was less than 0.05).

require the identification of additional patients with rare atypical rearrangements in this region.

METHODS SUMMARY

Underweight is defined in adults as BMI ≤ 18.5 . In individuals younger than 18 years of age, it is defined as a Z-score ≤ -2 .

Statistics. Two-tailed Fisher's exact test was used to compare frequencies of the rearrangement in patients and controls. Z-scores were computed for all data using gender-, age- and geographically-matched reference populations. One-tailed Student's *t*-test was performed to test BMI, height, weight and head circumference in duplication carriers for Z-scores of less than zero. We used Kruskal–Wallis tests for differences in gene expression patterns. *P*-values were adjusted using a Bonferroni correction, considering the number of pairwise comparisons; the resulting adjusted *P*-value was less than 0.05. The relative risk of being underweight was calculated as the ratio of the fraction of underweight individuals among duplication carriers versus our control group.

Discovery of CNVs. Carriers of 16p11.2 duplication and deletion were identified through various procedures: (1) comparative genomic hybridization with Agilent 44K, 60K, 105K, 180K, 244K arrays; (2) Illumina Human317, Human370, HumanHap550, Human610 and 1M BeadChips; (3) Affymetrix 6.0, 500K genotyping arrays; (4) quantitative multiplex PCR of short fluorescent fragments (QMPFSF); (5) fluorescent *in situ* hybridization (FISH); (6) MLPA. CNV analyses of GWAS data were carried out using cnvHap, a moving-window average-intensity procedure, a Gaussian mixture model, circular binary segmentation, QuantiSNP, PennCNV, BeadStudio GT module and Birdseed. At least two independent algorithms were used for each cohort.

Expression analyses. Lymphoblastoid cell lines were established from carriers and controls. SYBR Green quantitative PCR was performed to assess relative expression of genes.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 February; accepted 29 July 2011.

Published online 31 August 2011.

- Berrington de Gonzalez, A. *et al.* Body-mass index and mortality among 1.46 million white adults. *N. Engl. J. Med.* **363**, 2211–2219 (2010).
- Flegal, K. M., Graubard, B. I., Williamson, D. F. & Gail, M. H. Cause-specific excess deaths associated with underweight, overweight, and obesity. *J. Am. Med. Assoc.* **298**, 2028–2037 (2007).
- Olsen, E. M. *et al.* Failure to thrive: the prevalence and concurrence of anthropometric criteria in a general infant population. *Arch. Dis. Child.* **92**, 109–114 (2007).
- Corbett, S. S. & Drewett, R. F. To what extent is failure to thrive in infancy associated with poorer cognitive development? A review and meta-analysis. *J. Child Psychol. Psychiatry* **45**, 641–654 (2004).
- Rudolf, M. C. & Logan, S. What is the long term outcome for children who fail to thrive? A systematic review. *Arch. Dis. Child.* **90**, 925–931 (2005).
- Bravender, T. *et al.* Classification of eating disturbance in children and adolescents: proposed changes for the DSM-V. *Eur. Eat. Disord. Rev.* **18**, 79–89 (2010).
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR* 39–134 and 583–596 (Am. Psychiatric Assoc., 2000).
- Scherag, S., Hebebrand, J. & Hinney, A. Eating disorders: the current status of molecular genetic research. *Eur. Child Adolesc. Psychiatry* **19**, 211–226 (2010).
- Wang, K. *et al.* A genome-wide association study on common SNPs and rare CNVs in anorexia nervosa. *Mol. Psychiatry* doi:10.1038/mp.2010.107 (2010).
- Walters, R. G. *et al.* A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* **463**, 671–675 (2010).
- Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477–488 (2008).
- McCarthy, S. E. *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genet.* **41**, 1223–1227 (2009).
- Weiss, L. A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667–675 (2008).
- Crespi, B., Stead, P. & Elliot, M. Evolution in health and medicine Sackler colloquium: comparative genomics of autism and schizophrenia. *Proc. Natl Acad. Sci. USA* **107** (Suppl 1), 1736–1741 (2010).
- Bochukova, E. G. *et al.* Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature* **463**, 666–670 (2010).
- Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* **8**, 6 (2008).
- Sabatti, C. *et al.* Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature Genet.* **41**, 35–46 (2009).
- Nelis, M. *et al.* Genetic structure of Europeans: a view from the north-east. *PLoS ONE* **4**, e5472 (2009).
- Girirajan, S. & Eichler, E. E. Phenotypic variability and genetic susceptibility to genomic disorders. *Hum. Mol. Genet.* **19**, R176–R187 (2010).
- Pramyothin, P. & Khaothiar, L. Metabolic syndrome with the atypical antipsychotics. *Curr. Opin. Endocrinol. Diabetes Obes.* **17**, 460–466 (2010).
- Shinawi, M. *et al.* Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioral problems, dysmorphism, epilepsy, and abnormal head size. *J. Med. Genet.* **5**, 332–341 (2009).
- Merla, G. *et al.* Submicroscopic deletion in patients with Williams–Beuren syndrome influences expression levels of the nonhemizygous flanking genes. *Am. J. Hum. Genet.* **79**, 332–341 (2006).
- Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848–853 (2007).
- Henrichsen, C. N. *et al.* Segmental copy number variation shapes tissue transcriptomes. *Nature Genet.* **41**, 424–429 (2009).
- Ricard, G. *et al.* Phenotypic consequences of copy number variation: insights from Smith–Magenis and Potocki–Lupski syndrome mouse models. *PLoS Biol.* **8**, e1000543 (2010).
- Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genet.* **41**, 25–34 (2009).
- Courchesne, E., Carper, R. & Akshoomoff, N. Evidence of brain overgrowth in the first year of life in autism. *J. Am. Med. Assoc.* **290**, 337–344 (2003).
- Baxter, P. S., Rigby, A. S., Rotsaert, M. H. & Wright, I. Acquired microcephaly: causes, patterns, motor and IQ effects, and associated growth changes. *Pediatrics* **124**, 590–595 (2009).
- Li, Y., Dai, Q., Jackson, J. C. & Zhang, J. Overweight is associated with decreased cognitive functioning among school-age children and adolescents. *Obesity (Silver Spring)* **16**, 1809–1815 (2008).
- de Onis, M., Blossner, M., Borghi, E., Frongillo, E. A. & Morris, R. Estimates of global prevalence of childhood underweight in 1990 and 2015. *J. Am. Med. Assoc.* **291**, 2600–2606 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Vital-IT high-performance computing centre of the Swiss Institute of Bioinformatics. S.J. is recipient of a bourse de relève académique de la Faculté de Biologie et Médecine de l'Université de Lausanne. This work was supported by the Leenaards Foundation Prize (S.J., D.M. and A.Reymond), the Jérôme Lejeune Foundation (A.Reymond), the Telethon Action Suisse Foundation (A.Reymond), the Swiss National Science Foundation (A.Reymond, J.S.B., S.B. and S.E.A.), a Swiss National Science Foundation Sinergia grant (S.J., D.M., S.B., J.S.B. and A.Reymond), the European Commission anEUUploidy Integrated Project grant 037627 (A.Reymond, S.B., X.E., H.G.B. and S.E.A.), the Ludwig Institute for Cancer Research (A.V.), the Swiss Institute of Bioinformatics (S.B. and Z.K.), an Imperial College Department of Medicine PhD studentship (J.S.E.-S.M.), the Comprehensive Biomedical Research Centre, Imperial College Healthcare NHS Trust, and the National Institute for Health Research (P.E.), the Wellcome Trust and the Medical Research Council (A.I.F.B. and P.F.), the Instituto de Salud Carlos III (ISCIII)-FIS, the German Mental Retardation Network funded through a grant of the German Federal Ministry of Education and Research (NGFNplus 01GS08160) (A.Reis), European Union-FEDER (PI081714, PS09/01778) (F.F.A., M.G. and X.E.), SAF2008-02278 (C.R.), the Belgian National Fund for Scientific Research, Flanders (N.V.A. and R.F.K.), the Dutch Organisation for Health Research and Development (ZON-MW grant 917-86-319) and Hersenstichting Nederland (B.B.A.D.V.), grant 81000346 from the Chinese National Natural Science Foundation (Y.G.Y.), the Simons Foundation Autism Research Initiative, Autism Speaks and NIH grant GM061354 (J.F.G.), and Oesterreichische Nationalbank (OENB) grant 13059 (A.K.-B.). Y.S. holds a Young Investigator Award from the Children's Tumor Foundation and a Catalyst Award from Harvard Medical School. B.L.W. holds a Fudan Scholar Research Award from Fudan University, a grant from Chinese National '973' project on Population and Health (2010CB529601) and a grant from the Science and Technology Council of Shanghai (09JC1402400). E.R.S. and S.L., recipients of the Michael Smith Foundation for Health Research scholar award, acknowledge the CIHR MOP 74502 operational grant. The Estonian Genome Center of the University of Tartu (EGCUT) received support from the EU Centre of Excellence in Genomics and FP7 grants 201413 and 245536, and from Estonian Government SF0180142s08, SF0180026s09 and SF0180027s10 (A.M., K.M. and A.K.). D.S. thanks the Direction Générale de l'Organisation des Soins from the French Ministry of Health for their support in the development of several array-CGH platforms, and the Centres Labellisés Anomalies du Développement en France. The Helmholtz Zentrum Munich and the State of Bavaria financed the KORA study, also supported by the German National Genome Research Network (NGFN-2 and NGFNplus: 01GS0823), the German Federal Ministry of Education and Research (BMBF), and the Munich Center of Health Sciences (MC Health, LMUinnovativ). CIBERON and CIBERESP are initiatives of ISCIII (Spain). S.W.S. holds the GlaxoSmithKline-Canadian Institutes of Health Chair in Genetics and Genomics at the University of Toronto and the Hospital for Sick Children, and is supported by Genome Canada and the McLaughlin Centre. Funding for deCODE came in part from NIH grant MH071425 (K.S.), EU grant HEALTH-2007-2.2.1-10-223423 (Project PsychCNV) and EU grant IMI-JU-NewMeds. NFBC1966 received financial support from the Academy of Finland (project grants 104781, 120315, 129269, 1114194, Center of Excellence in Complex Disease Genetics and SALVE), University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), the European Commission (EURO-BLCS, Framework 5 award QLGI-CT-2000-01643), NHLBI grant 5R01HL087679-02 through the STAMPEED program (1RL1MH083268-01), NIH/NIMH (5R01MH63706:02), ENGAGE project and grant agreement HEALTH-F4-2007-201413, and the Medical Research Council, UK (G0500539, G0600705, PrevMetSyn/SALVE). The DNA extractions, sample quality controls, biobank up-keeping and aliquotting was performed in the National Public Health Institute, Biomedicum Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki. We thank M. Hass, Z. Jaros, M. Jussila, M. Koiranen,

P. Rantakallio, M. C. Rudolf, V. Soo, O. Tornwall, S. Vaara, T. Ylitalo and the French DHOS national CGH network for their help, as well as all participating patients and clinicians. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions S.J., A.Reymond, P.F. and J.S.B. wrote the manuscript with contributions from F.Z., L.H., R.G.W., N.D.B., Z.K., A.I.F.B. and A.V. L.H., A.V. and A.Reymond produced and analyzed the expression data. Z.K., A.V., R.G.W. and N.D.B. conducted the statistical analyses, guided by S.J., A.Reymond, P.F. and J.S.B. S.J., A.Reymond, F.Z., L.H., D.M., Y.S., G.T., M.B., S.B., D.C., N.d.L., B.B.A.d.V., B.A.F., F.F.A., M.G., A.G., J.H., A.K., C.L.C., K.M., O.S.P. D.S., M.M.V.H., S.V.G., A.T.V.-v.S., F.W., B.-L.W., Y.Y., J.A., X.E., J.F.G., A.M., S.W.S., K.S., U.T., A.I.F.B., J.S.B., P.F. and all other authors phenotyped and/or genotyped patients and/or individuals of the general population. S.J., A.Reymond and J.S.B. designed the study. All authors commented on and approved the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. J.S.B. and P.F. are equally contributing senior authors and F.Z., L.H., R.G.W. and Z.K. are equally contributing second authors. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.S.B. (Jacques.Beckmann@chuv.ch).

Sébastien Jacquemont^{1*}, Alexandre Reymond^{2*}, Flore Zufferey¹, Louise Harewood², Robin G. Walters³, Zoltán Kutalik^{4,5}, Danielle Martinet¹, Yiping Shen^{6,7}, Armand Valsesia^{4,5,8}, Noam D. Beckmann¹, Gudmar Thorleifsson⁹, Marco Belfiore¹, Sonia Bouquillon¹⁰, Dominique Campion¹¹, Nicole de Leeuw¹², Bert B. A. de Vries¹², Tõnu Esko^{3,14}, Bridget A. Fernandez¹⁵, Fernando Fernández-Aranda¹⁶, José Manuel Fernández-Real¹⁷, Mónica Gratacós¹⁸, Audrey Guilmatre¹¹, Juliane Hoyer¹⁹, Marjo-Riitta Jarvelin^{20,21,22}, R. Frank Kooy²³, Ants Kurg¹³, Cédric Le Caignec²⁴, Katrin Männik¹³, Orah S. Platt⁶, Damien Sanlaville²⁵, Mieke M. Van Haelst^{3,26}, Sergi Villatoro Gomez¹⁸, Faida Walha², Bai-lin Wu^{6,27}, Yongguo Yu^{6,28}, Azzedine Aboura²⁹, Marie-Claude Addor¹, Yves Alembik³⁰, Stylianos E. Antonarakis³¹, Benoît Arveiler^{32,33}, Magalie Barth³⁴, Nathalie Bednarek³⁵, Frédérique Béna³¹, Sven Bergmann^{4,5}, Mylène Beri³⁶, Laura Bernardini³⁷, Bettina Blaumeiser²³, Dominique Bonneau³⁴, Armand Bottani³¹, Odile Boute³⁸, Han G. Brunner¹², Dorothee Caillay³³, Patrick Callier³⁹, Jean Chiesa⁴⁰, Jacqueline Chrat², Lachlan Coin³, Charles Coutton⁴¹, Jean-Marie Cuisset⁴², Jean-Christophe Cuvellier⁴², Albert David²⁴, Bénédicte de Freminville⁴³, Bruno Delobel⁴⁴, Marie-Ange Delrue^{32,33}, Bénédicte Demeer⁴⁵, Dominique Descamps⁴⁶, Gérard Didelot², Klaus Dieterich⁴⁷, Vittoria Disciglio^{48,110}, Martine Doco-Fenzy⁴⁹, Séverine Drunat²⁹, Bénédicte Duban-Bedu⁴⁴, Christèle Dubourg^{50,51}, Julia S. El-Sayed Moustafa², Paul Elliott^{52,53}, Brigitte H. W. Faas¹², Laurence Faivre⁵⁴, Anne Faudet⁵⁵, Florence Fellmann¹, Alessandra Ferrarini⁵⁶, Richard Fisher⁵⁶, Elisabeth Flori⁵⁰, Lukas Forer⁵⁷, Dominique Gaillard⁴⁹, Marion Gerard²⁹, Christian Gieger⁵⁸, Stefania Gimelli⁵¹, Giorgio Gimelli⁵⁹, Hans J. Grabe⁶⁰, Agnès Guichet³⁴, Olivier Guillin¹¹, Anna-Liisa Hartikainen⁶¹, Delphine Heron^{62,63}, Loyse Hippolyte¹, Muriel Holder³⁸, Georg Homuth⁶⁴, Bertrand Isidor²⁴, Sylvie Jaillard^{50,65}, Zdenek Jaros⁶⁶, Susana Jimenez-Murcia¹⁶, Geraldine Joly Helas⁶⁷, Philippe Jonveaux⁶⁸, Satu Kaksanen⁶⁸, Boris Keren⁵⁵, Anita Kloss-Brandstätter⁵⁷, Nine V. A. M. Knoers^{26,69}, David A. Kooren¹², Peter M. Kroisel⁷⁰, Florian Kronenberg⁵⁷, Audrey Labalme²⁵, Emilie Landais⁴⁹, Elisabetta Lapi⁷¹, Valérie Layet⁷², Solenn Legallie¹¹, Bruno Leheup⁷³, Barbara Leube⁷⁴, Suzanne Lewis⁷⁵, Josette Lucas⁶⁵, Kay D. MacDermot⁷⁶, Pall Magnusson⁷⁷, Christian Marshall⁷⁸, Michèle Mathieu-Dramard⁴⁵, Mark I. McCarthy^{79,80,81}, Thomas Meitinger^{82,111}, Maria Antonietta Mencarelli⁴⁸, Giuseppe Merla⁸³, Alexandre Moerman³, Vincent Mosser⁸⁴, Fanny Morice-Picard^{32,33}, Mafalda Mucciolo⁴⁸, Matthias Nauck⁸⁵, Ndeye Coumba Ndiaye⁸⁶, Ann Nordgren⁸⁷, Laurent Pasquier⁸⁸, Florence Petit³⁸, Rolf Pfundt¹², Ghislaine Plessis⁸⁹, Evica Rajcan-Separovic⁹⁰, Gian Paolo Ramelli⁹¹, Anita Rauch⁹², Roberto Ravazzolo⁹³, Andre Reis¹⁹, Alessandra Renieri⁴⁸, Cristobal Richart⁹⁴, Janina S. Ried⁵⁸, Claudine Rieubland⁹⁵, Wendy Roberts⁹⁶, Katharina M. Roetzer⁷⁰, Caroline Rooryck^{32,33}, Massimiliano Rossi²⁵, Evald Saemundsen⁹⁷, Véronique Satre⁴¹, Claudia Schurmann⁶⁴, Engilbert Sigurdsson⁹⁸, Dimitri J. Stavropoulos⁹⁹, Hreinn Stefansson⁹, Carola Tengström¹⁰⁰, Unnur Thorsteinsdóttir^{9,101}, Francisco J. Tinahones¹⁰², Renaud Touraine⁴³, Louis Vallée⁴², Ellen van Binsbergen²⁶, Nathalie Van der Aa²³, Catherine Vincent-Delorme¹⁰³, Sophie Visvikis-Siest⁹⁶, Peter Vollenweider¹⁰⁴, Henry Völzke¹⁰⁵, Anneke T. Vulto-van Silfhout¹², Gérard Waeber¹⁰⁴, Carina Wallgren-Pettersson¹⁰⁶, Robert M. Witwicki², Simon Zwiolinski⁵⁶, Joris Andrieux¹⁰, Xavier Estivill¹⁸, James F. Gusella⁷, Omar Gustafsson^{9,107}, Andres Metspalu^{13,14}, Stephen W. Scherer¹⁰⁸, Kari Stefansson⁹, Alexandra I. F. Blakemore³, Jacques S. Beckmann^{1,4} & Philippe Froguel^{3,109}

¹Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland. ²Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. ³Department of Genomics of Common Disease, Imperial College London, London W12 0NN, UK. ⁴Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland. ⁵Swiss Institute of Bioinformatics, University of Lausanne, 1015 Lausanne, Switzerland. ⁶Laboratory Medicine, Children's Hospital Boston, Boston, Massachusetts 02115, USA. ⁷Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁸Ludwig Institute for Cancer Research, University of Lausanne, 1015 Lausanne, Switzerland. ⁹deCODE Genetics, Sturlugata 8, IS-101 Reykjavik, Iceland. ¹⁰Laboratoire de Génétique Médicale, Hôpital Jeanne de Flandre, CHRU de Lille, 59000 Lille, France. ¹¹INSERM U614, University of Rouen and Centre Hospitalier du Rouvray, 76000 Rouen, France. ¹²Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, 6500 HB

Nijmegen, The Netherlands. ¹³Institute of Molecular and Cell Biology, University of Tartu, 51010 Tartu, Estonia. ¹⁴Estonian Genome Center, University of Tartu, 51010 Tartu, Estonia. ¹⁵Discipline of Genetics and Medicine, Memorial University of Newfoundland, St John's A1B 3V6, Newfoundland, Canada. ¹⁶Department of Psychiatry, University Hospital of Bellvitge-IDIBELL, Ciber Fisiopatología Obesidad y Nutrición (CIBEROBN), 08907 Barcelona, Spain. ¹⁷Department of Diabetes, Endocrinology, and Nutrition, Hospital Universitari de Girona Dr. Josep Trueta, Institut d'Investigació Biomèdica de Girona, Ciber Fisiopatología Obesidad y Nutrición (CIBEROBN), Instituto Salud Carlos III, 17007 Girona, Spain. ¹⁸Genes and Disease Program, Center for Genomic Regulation (CRG-UPF), CIBER en Epidemiología y Salud Pública (CIBERESP), 08003 Barcelona, Catalonia, Spain. ¹⁹Institute of Human Genetics, Friedrich-Alexander University Erlangen-Nuremberg, 91054 Erlangen, Germany. ²⁰Department of Epidemiology and Biostatistics, Imperial College London, School of Public Health, London W2 1PG, UK. ²¹Department of Child and Adolescent Health, National Institute for Health and Welfare, Box 310, 90101 Oulu, Finland. ²²Institute of Health Sciences, University of Oulu, and Biocenter Oulu, University of Oulu, Box 5000, 90014 University of Oulu, Finland. ²³Department of Medical Genetics, University and University Hospital Antwerp, B-2650 Antwerp, Belgium. ²⁴Service de Génétique Médicale, CHU Nantes, 44093 Nantes, France. ²⁵Service de Cytogénétique Constitutionnelle, Hospices Civils de Lyon, CHU de Lyon and Neuroscience Research Center, TIGER team, UCLB1, Lyon, F-69000, France. ²⁶Department of Medical Genetics, University Medical Center, 3584 EA Utrecht, The Netherlands. ²⁷Children's Hospital and Institutes of Biomedical Science, Fudan University, 200032 Shanghai, China. ²⁸Shanghai Children's Medical Center, 200127 Shanghai, China. ²⁹Department of Genetics, APHP-Robert DEBRE University Hospital, 75019 Paris, France. ³⁰Service de Génétique Médicale, CHU Strasbourg, Hôpital de Hautepierre, 67000 Strasbourg, France. ³¹Service of Genetic Medicine, University Hospitals of Geneva, 1205 Geneva, Switzerland. ³²Laboratoire Maladies Rares-Génétique et Métabolisme, Université Bordeaux 2, 33076 Bordeaux, France. ³³Service de Génétique Médicale, CHU de Bordeaux, 33076 Bordeaux, France. ³⁴Service de Génétique, CHU Angers, 49933 Angers, France. ³⁵Service Pédiatrie, CHU Hôpital Alix de Champagne, 51100 Reims, France. ³⁶Laboratoire de génétique, CHU Nancy, 54511 Vandœuvre les Nancy, France. ³⁷Mendel Laboratory, IRCCS Casa Sollievo della Sofferenza Hospital, 71013 San Giovanni Rotondo, Italy. ³⁸Service de Génétique Clinique, Hôpital Jeanne de Flandre, CHRU de Lille, 59037 Lille, France. ³⁹Laboratoire de Cytogénétique, CHU le Bocage, 21070 Dijon, France. ⁴⁰Laboratoire de Cytogénétique, CHU Caremeau, 30029 Nîmes, France. ⁴¹Laboratoire de Génétique Chromosomique, CHU de Grenoble, BP 217, 38043 Grenoble, France. ⁴²Service de Neurologie pédiatrique, Hôpital Roger Salengro, CHRU de Lille, 59037 Lille, France. ⁴³CHU-Hôpital Nord, Service de Génétique, CHU Saint Etienne, 42055 St Etienne, France. ⁴⁴Centre de Génétique Chromosomique, Hôpital Saint-Vincent de Paul, GHICL, 59160 Lille, France. ⁴⁵Service de Génétique Médicale, CHRU Amiens, 80000 Amiens, France. ⁴⁶Centre hospitalier de Béthune, 62408 Bethune, France. ⁴⁷Service de Génétique Clinique, CHU Grenoble, 38043 Grenoble, France. ⁴⁸Medical Genetics, Department of Biotechnology, University of Siena, 53100 Siena, Italy. ⁴⁹Service de Génétique, HMB, CHU REIMS, IFR 53, 51092 Reims, France. ⁵⁰UMR 6061 CNRS, IFR 140 GFAS, Université de Rennes 1, 35043 Rennes, France. ⁵¹Service de Génétique Moléculaire, CHU, 35033 Rennes, France. ⁵²Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK. ⁵³MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK. ⁵⁴Centre de Génétique, CHU Dijon, 21079 Dijon, France. ⁵⁵Département de Génétique et Cytogénétique, APHP-GH Pitié-Salpêtrière, 75013 Paris, France. ⁵⁶Institute of Human Genetics, International Centre for Life, Newcastle Upon Tyne NE1 4EP, UK. ⁵⁷Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, 6020 Innsbruck, Austria. ⁵⁸Institute of Genetic Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health, 85764 Neuherberg, Germany. ⁵⁹Laboratorio di Citogenetica, G. Gaslini Institute, 16147 Genova, Italy. ⁶⁰Department of Psychiatry and Psychotherapy, Ernst-Moritz-Arndt University Greifswald, 17475 Greifswald and D-18437 Stralsund, Germany. ⁶¹Institute of Clinical Medicine, Department of Obstetrics and Gynecology, University of Oulu, 90570 Oulu, Finland. ⁶²Embryologie AP-HP, Université Pierre et Marie Curie, 75005 Paris, France. ⁶³Département de Génétique Cytogénétique, Hôpital Pitié-Salpêtrière, 75013 Paris, France. ⁶⁴Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt University Greifswald, D-17487 Greifswald, Germany. ⁶⁵Laboratoire de Cytogénétique et Biologie Cellulaire, CHU Rennes, 35033 Rennes, France. ⁶⁶Abteilung für Kinder und Jugendheilkunde, Landeskindernklinik Waldviertel Zwettl, 3910 Zwettl, Austria. ⁶⁷Department of Genetics, CHU Rouen, 76183 Rouen, France. ⁶⁸The Habilitation Unit of Folkhalsan, Folkhalsan, SF 00250 Helsinki, Finland. ⁶⁹Department of Human Genetics, University Medical Center, 3584 EA Utrecht, The Netherlands. ⁷⁰Institute of Human Genetics, Medical University of Graz, 8010 Graz, Austria. ⁷¹Medical Genetics Unit, Children's Hospital Anna Meyer, 50139 Firenze, Italy. ⁷²Unité de Génétique, Groupe Hospitalier du Havre, 76600 le Havre, France. ⁷³Service de Médecine Infantile III et Génétique Clinique, CHU-Nancy et PRES de l'Université de Lorraine UHP Nancy, 54511 Vandœuvre les Nancy, France. ⁷⁴Institute of Human Genetics and Anthropology, Heinrich-Heine University Hospital Duesseldorf, D-40001 Duesseldorf, Germany. ⁷⁵Department of Medical Genetics, The University of British Columbia and Child and Family Research Institute, Vancouver V6H 3N1, Canada. ⁷⁶North West Thames Regional Genetics Service, Northwick Park & St Marks Hospital, Harrow HA1 3UJ, UK. ⁷⁷Child and Adolescent Psychiatry, Landspítali University Hospital, IS-105 Reykjavik, Iceland. ⁷⁸The Centre for Applied Genomics, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ⁷⁹Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Headington, Oxford OX3 7LJ, UK. ⁸⁰Wellcome Trust Centre for Human Genetics, University of Oxford, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK. ⁸¹Oxford NIHR Biomedical Research Centre, Churchill Hospital, Old Road, Headington, Oxford OX3 7LJ, UK. ⁸²Institute of Human Genetics, Helmholtz Center Munich, German Research Center for Environmental Health and Institute of Human Genetics, Technical University Munich, 85764 Neuherberg, Germany. ⁸³Medical Genetics Unit, IRCCS Casa Sollievo della Sofferenza Hospital, 71013 San Giovanni

Rotondo, Italy. ⁸⁴Genetics, GlaxoSmithKline R&D, 720 Swedeland Road, King of Prussia, Pennsylvania 19406, USA. ⁸⁵Institute of Clinical Chemistry and Laboratory Medicine, Ernst-Moritz-Arndt University Greifswald, D-17475 Greifswald, Germany. ⁸⁶Cardiovascular Genetics Research Unit, EA4373, Université Henri Poincaré, 54000 Nancy, France. ⁸⁷Department of Molecular Medicine and Surgery, Karolinska Institutet, 171 76 Stockholm, Sweden. ⁸⁸Service de Génétique-CLAD Ouest, 35203 Rennes cedex 2, France. ⁸⁹Service de Génétique, CHU Clémenceau, 14033 Caen, France. ⁹⁰Department of Pathology, University of British Columbia and Child and Family Research Institute, Vancouver, British Columbia V5Z 4H4, Canada. ⁹¹Division of Pediatrics, Ospedale San Giovanni, 6500 Bellinzona, Switzerland. ⁹²Institute of Medical Genetics, University of Zurich, 8603 Schwerzenbach, Switzerland. ⁹³Department of Pediatrics and CEBR, G. Gaslini Institute, University of Genova, 16126 Genova, Italy. ⁹⁴Department of Internal Medicine, University Hospital Juan XXIII, Universitat Rovira y Virgili, Ciber Fisiopatología Obesidad y Nutrición (CIBEROBN), Instituto Salud Carlos III, 43005 Tarragona, Spain. ⁹⁵Division of Human Genetics, Department of Paediatrics, Inselspital, University of Bern, 3010 Bern, Switzerland. ⁹⁶The Autism Research Unit, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ⁹⁷State Diagnostic and Counseling Center, 200 Kopavogur, Iceland. ⁹⁸University of Iceland and Landspítali University Hospital, 101 Reykjavik,

Iceland. ⁹⁹Department of Pediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ¹⁰⁰Genetic Services, Rinnekoti Research Foundation, Kumputie 1, SF-02980 Espoo, Finland. ¹⁰¹Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland. ¹⁰²Department of Endocrinology and Nutrition, Clinic Hospital of Virgen de la Victoria, Ciber Fisiopatología Obesidad y Nutrición (CIBEROBN), Instituto Salud Carlos III, 29010 Malaga, Spain. ¹⁰³Centre de Maladies Rares, Anomalies du Développement Nord de France, CH Arras-CHRU Lille, 59000 Lille, France. ¹⁰⁴Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland. ¹⁰⁵Institute for Community Medicine, Ernst-Moritz-Arndt University Greifswald, D-17475 Greifswald, Germany. ¹⁰⁶Department of Medical Genetics, Haartman Institute, University of Helsinki and Folkhälsan Institute of Genetics, 00251 Helsinki, Finland. ¹⁰⁷Department of Psychiatry, Oslo University Hospital, N-0407 Oslo, Norway. ¹⁰⁸The Hospital for Sick Children, University of Toronto, Toronto, Ontario M5G 1L7, Canada. ¹⁰⁹CNRS 8090-Institute of Biology, Pasteur Institute, 59800 Lille, France. ¹¹⁰UOC Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy. ¹¹¹Institute of Human Genetics, Technische Universität München, Klinikum rechts der Isar, 81675 München, Germany.

*These authors contributed equally to this work.

METHODS

Study cohorts. For the description of these cohorts, refer to Supplementary Information.

CNV detection. Cases ascertained for intellectual disabilities and developmental delay were identified through standard medical diagnostic procedures. CNV analyses of GWAS data were variously carried out using *cnvHap*³¹; a moving-window average-intensity procedure; a Gaussian mixture model (Valsesia *et al.*, submitted); circular binary segmentation^{32,33}; *QuantiSNP*³⁴; *PennCNV*³⁵; *BeadStudio* GT module (Illumina Inc.); and *Birdseed*³⁶ (see below). At least two independent algorithms were used for each cohort.

Patients referred for intellectual disabilities and developmental delay. All diagnostic procedures (CGH, quantitative PCR and/or quantitative multiplex PCR of short fluorescent fragments) were carried out according to the relevant guidelines of good clinical laboratory practice for the respective countries. All rearrangements in probands were confirmed by a second independent method and karyotyping was performed in all cases to exclude a complex rearrangement. **Northern Finland 1966 birth cohort (NFBC).** CNV calling has been previously described¹⁰. In brief, data were normalized using *Illumina* *BeadStudio*, then GC effects on ratios were removed by regressing on GC and GC2, and wave effects were removed by fitting a Loess function³⁷. CNV analysis was done using *cnvHap*³¹. All called 16p11.2 duplications were validated by direct analysis of log₂ ratios. Data for each probe were normalized by first subtracting the median value across all samples (so that the distribution of ratios for each probe was centred on zero), and then dividing by the variance across all samples (to correct for variation in the sensitivity of different probes to copy-number variation). All CNV calls were confirmed by MLPA.

deCODE genetics. *Illumina* Human317, Human370, HumanHap550, Human610 and 1M *BeadChips* were used for CNV analysis. *BeadStudio* (version 2.0) was used to call genotypes, normalize the signal intensity data and establish the log R ratio (LRR) and B allele frequency (BAF) at every SNP according to standard *Illumina* protocols. All samples passed a standard SNP-based quality control procedure with a SNP call rate greater than 0.97. *PennCNV*³⁵, a free, open-source tool, was used for detection of CNVs. The input data for *PennCNV* are LRR, a normalized measure of the total signal intensity for the two alleles of the SNP, and BAF, a normalized measure of the allelic intensity ratio of the two alleles. These values are derived with the help of control genotype clusters (*HapMap* samples), using the *Illumina* *BeadStudio* software. *PennCNV* employs a hidden Markov model to analyse the LRR and BAF values across the genome. CNV calls are made on the basis of the probability of a given copy state at the current marker, as well as on the probability of observing a copy-state change from the previous marker to the current one. *PennCNV* uses a built-in correction model for GC content³⁸.

Cohorte Lausannoise (CoLaus). Data normalization and CNV calling have been previously described¹⁰. Data normalization included allelic cross-talk calibration^{39,40}, intensity summarization using robust median average, and correction for any PCR amplification bias. Wave effects were corrected by fitting a Loess function³⁷. CNV calling was done using a Gaussian mixture model (Valsesia *et al.*, submitted) that fits four components (deletion, copy-neutral, one additional copy and two additional copies) to copy-number ratios. The final copy number at each probe location is determined as the expected (dosage) copy number. The method has been validated by comparing test data sets with results from the *CNAT*⁴¹ and *CBS*^{32,33} algorithms, and by replicating a subset of CoLaus subjects on *Illumina* arrays. Only duplications found by both Gaussian mixture model and *CBS* were considered.

Estonian genome center of the University of Tartu (EGCUT). Genotypes were called by *BeadStudio* software GT module v3.1 or *GenomeStudio* GT v1.6 (*Illumina* Inc.). Values for LRR and BAF produced by *BeadStudio* were formatted for further CNV analysis and break-point mapping with Hidden-Markov-Model-based softwares *QuantiSNP* (ver.1.1)³⁴ and *PennCNV*⁴² or *CNVPartition* 2.4.4 (*Illumina* Inc.). All analyses were carried out using the recommended settings, except changing EMitters to 25 and L to 1,000,000 in *QuantiSNP*. For *PennCNV*, the Estonian-population-specific SNP allele frequency data was used. All detected duplications were confirmed by quantitative PCR.

Study of health in Pomerania (SHIP). Raw intensities were normalized using *Affymetrix* power tools (*Affymetrix*); CNV analysis was done using *Birdseye* from the *Birdsuite* software package³⁶ and *PennCNV*³⁵. *PennCNV* predictions with confidence scores less than 10 were removed. *Birdsuite* predictions were filtered as in ref. 15: CNVs were kept if their linkage disequilibrium (LOD) score was >10, length >1 kb, number of probes ≥5 and size per number of probes <10,000.

Kooperative Gesundheitsforschung in der Region Augsburg (KORA) F3 and F4. Genotyping for KORA F3 was performed using the *Affymetrix* 500K array set, consisting of two chips (StyI and NspI). The KORA F4 samples were genotyped with the *Affymetrix* human SNP array 6.0. For both studies, genomic DNA from blood samples was used for analysis. Hybridization of genomic DNA was done in accordance with the manufacturer's standard recommendations. Genotyping was

done in the Genome Analysis Centre of the Helmholtz Centre Munich. Genotypes were determined using BRLMM clustering algorithm (*Affymetrix* 500K array set) and *Birdseed2* clustering algorithm (*Affymetrix* array 6.0). For quality control purposes, we applied a positive control and a negative control DNA every 48 samples (KORA F3) or 96 samples (KORA F4). On the chip level, only subjects with overall genotyping efficiencies of at least 93% were included. In addition, the called gender had to agree with the gender in the KORA study database. After exclusions, 1,644 individuals remained in KORA F3 and 1,814 in KORA F4 for further analysis.

MLPA analysis. We used MLPA to determine changes in the copy number of a region of about 2 Mb on chromosome 16p11.2. Briefly, we designed, using hg18, nine probes within the targeted region, one control probe outside the rearranged region and seven control probes targeting unique position in the genome (Supplementary Table 10). Assays were performed with MRC-Holland reagents according to the manufacturer's protocol⁴³. The analysis of the amplification products was performed by capillary electrophoresis in the DNA Analyser 3730XL and using the *GeneMapper* software v3.7 (*Applied Biosystems*). The calculations were performed independently for each experiment: we first normalized the MLPA data to minimize the amount of experimental variation, summing all signal values of each control probe for each sample, and then dividing each signal value of each sample by the sum. The normalized signal values were compared to signal values from all other samples in the same experiment, dividing the normalized signal values by the average calculated from all the samples in the same experiment. The product of this calculation is termed dosage quotient (DQ). A DQ value of less than 0.65 or more than 1.25 was considered as copy-number loss or gain, respectively, as previously described^{44–46}.

Custom array-CGH for the short arm of chromosome 16. DNA samples were labelled with Cy3 and cohybridized to custom-made Nimblegen arrays with Cy5-labelled DNA from the CEPH cell line GM12042. These arrays contained 71,000 probes spread across the short arm of chromosome 16 from 22.0 Mb to 32.7 Mb (at a median space of 45 bp between 27.5 Mb and 31.0 Mb), and 1,000 control probes situated in invariable regions of the X chromosome. DNA labelling, hybridization and washing were performed according to Nimblegen protocols. Scanning was performed using an *Agilent* G2565BA microarray scanner. Image processing, quality control and data extraction were performed using the *Nimblescan* software v2.5.

Defining underweight. Underweight was defined throughout the study as BMI ≤ 18.5 kg per m² in adults and ≤ −2 s.d. in children^{30,47,48}.

Weight, height, BMI and head circumference Z-scores as a function of age. For paediatric cases, weight, height, BMI and head circumference Z-scores were determined for paediatric cases (0–18 years of age) using clinical growth charts specific to the country of origin. Children were ascertained from nine different countries. If charts were only available in percentiles, those measures were transformed into Z-scores using gender-, age- and geographically-matched reference populations (see Statistics).

For the USA and Canada, data from the Center for Disease Control and National Center for Health Statistics (CDC/NCHS) were used to calculate Z-scores⁴⁹.

For the French paediatric population, we used French national growth charts^{50,51}. For the Swiss paediatric population, we used Swiss national growth charts⁵². For Dutch participants, Dutch national growth charts were used⁵³. For Italian, German, Finnish and Austrian cases (*n* = 6), height, weight and BMI Z-scores were estimated using WHO growth charts⁵⁴.

To check for discrepancies generated by the use of different growth charts, height, weight and BMI Z-scores were recalculated using WHO growth charts for all cases under five years of age, regardless of origin (<http://www.who.int/childgrowth/standards/en/54>). Z-scores obtained using the WHO data were not significantly different. These growth standards, developed by the World Health Organization multicentre growth reference study, describe normal child growth from birth to 5 years under optimal environmental conditions. These standards can be applied to all children everywhere, regardless of ethnicity, socioeconomic status and type of feeding^{55,56}.

If necessary, percentile values were transformed to Z-scores by the inverse-normal density function. When growth charts were unavailable, we used reported LMS parameters (median (M), generalized coefficient of variation (S) and skewness (L)) to obtain Z-scores via the formula:

$$Z\text{-score} = \begin{cases} \frac{(X/M)^L - 1}{L \cdot S}, & L > 0 \\ \ln(X/M)/S, & L = 0 \end{cases}$$

in which X is the observed value.

In adults (>18 years of age), we estimated LMS parameters when these were unavailable from the available sex-, age- and origin-matched Swiss (CoLaus),

Estonian or French control populations. For cases identified from population-based cohorts, Z-scores were directly inferred from the cohort.

Gene expression. We established lymphoblastoid cell lines from deletion and duplication carriers, as well as from controls (Supplementary Table 12), by transforming peripheral blood mononuclear cells with Epstein–Barr virus. Patients and controls were enrolled after obtaining appropriate informed consent via the physicians in charge, and approval by the ethics committee of the University of Lausanne. More control cell lines were obtained from Coriell Institute for Medical Research (<http://www.coriell.org/>) (Supplementary Table 12). SYBR Green real-time quantitative PCR (RT–PCR) was performed as previously described^{22,27}. Briefly, 1 µg of total RNA from lymphoblastoid cell lines was converted to complementary DNA using Superscript VILO (Invitrogen) primed with a mixture of oligo(dT) and random hexamers. Oligos were designed using the PrimerExpress program (Applied Biosystems) with default parameters (Supplementary Table 11). Non-intron-spanning assays were tested for genomic contamination in standard \pm reverse transcriptase reactions. The amplification efficiency of each primer pair was tested in a cDNA dilution series, as previously described⁵⁸. A full list of genes mapping in the rearranged interval, and exclusion criteria, are presented in Supplementary Table 1. All RT–PCR reactions were performed in a 10-µl final volume and triplicates per sample. The setup in a 384-well plate format was performed using a Freedom EVO robot (TECAN) and assays were run in an ABI 7900 sequence detection system (Applied Biosystems) with the following amplification conditions: 50 °C for 2 min, 95 °C for 10 min, and 45 cycles of 95 °C 15 s, then 60 °C for 1 min. A final incubation of 95 °C for 15 s followed by 60 °C for 15 s was carried out to establish a dissociation curve. Each plate included the appropriate normalization genes to control for any variability between plate runs. Raw threshold cycles (Ct) values were obtained using SDS2.4 (Applied Biosystems). To calculate the normalized relative expression ratio of individuals carrying the CNV and of controls, we used Biogazelle qBase Plus software⁵⁹ including geNorm⁶⁰. This program identified appropriate normalization genes (*EEF1A1*, *RPL13*, *GUSB* and *TBP*) having a gene-stability measure of $M = 0.25$. We note that one gene, *LAT*, showed a very high expression profile in one of the duplication samples (DASYL, Supplementary Table 13), reaching a relative expression value of 27.3 (s.e.m. = 1.37), compared to an average expression for other duplications of 1.89 (s.e.m. = 0.51). We cannot exclude that this finding is genuine (and confirmed it in a second experiment), but it was removed from further analyses as an outlier to give a more accurate overview of expression profiles for these genes.

In silico analysis was performed to check for brain, and specifically hypothalamus, expression of genes in the rearranged 16p11.2 interval (Supplementary Table 1). This was done using Allen Brain Atlas Resources, available from <http://www.brain-map.org>.

Cases with major neurological signs. Major neurological signs were defined by moderate to severe hypotonia, hypertonia, ataxia, spasticity, hyperreflexia, hyporeflexia and/or extra-pyramidal signs, and by the presence of epilepsy.

Statistics. Student's *t*-test: one-tailed *t*-tests were performed to test whether duplication carriers have Z-score values lower than zero for BMI, height and weight. We found this analysis more suitable than linear regression analysis, correcting for confounding factors such as sex and age, because these anthropometric traits have a highly nonlinear dependence on these factors, as can be observed in control populations.

Kruskal–Wallis test: this was used to test differences in the gene expression pattern between deletion and duplication carriers and control individuals. Because expression values are not necessarily normally distributed, this test is more adequate than a classical one-way analysis of variance. To test pairwise differences, we computed the difference in mean group rank with its 95% confidence interval (as provided by the multcompare function in Matlab). Correction for multiple testing was done using a Bonferroni adjustment.

Multiple testing: we determined false-discovery-rate-based thresholds for association *P*-values for each phenotype, to correct for multiple testing. For each phenotype, we replaced the observed Z-scores with numbers randomly drawn from a standard normal distribution and performed the same *t*-tests for the same strata. The procedure was repeated 1,000 times. For various *P*-value thresholds, we asked how many tests would be declared significant for the null set on average (over the 1,000 random draws). The false discovery rate was estimated as the ratio of this number and the actual number obtained for the observed Z-scores. Thus, we controlled the dependence between nested tests.

Relative risk: among adults, we defined underweight as a BMI <18.5 (WHO criteria). The estimated relative risk is the ratio of the fraction of underweight individuals among duplication carriers versus our control group. The standard error of log(relative risk) and its significance were calculated as previously described⁶¹. In our control group (population-based cohorts), the frequency of

being underweight is 1.9% (38 males and 148 females out of 9,470). Owing to the fact that being underweight decreases with age in the general population, we resampled our control group to ensure precise age-matching.

31. Coin, L. J. *et al.* cnvHap: an integrative population and haplotype-based multiplatform model of SNPs and CNVs. *Nature Methods* **7**, 541–546 (2010).
32. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
33. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657–663 (2007).
34. Colella, S. *et al.* QuantiSNP: an objective Bayes Hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.* **35**, 2013–2025 (2007).
35. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).
36. Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature Genet.* **40**, 1253–1260 (2008).
37. Marioni, J. C. *et al.* Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol.* **8**, R228 (2007).
38. Diskin, S. J. *et al.* Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.* **36**, e126 (2008).
39. Bengtsson, H., Irizarry, R., Carvalho, B. & Speed, T. P. Estimation and assessment of raw copy numbers at the single locus level. *Bioinformatics* **24**, 759–767 (2008).
40. Bengtsson, H., Ray, A., Spellman, P. & Speed, T. P. A single-sample method for normalizing and combining full-resolution copy numbers from multiple platforms, labs and analysis methods. *Bioinformatics* **25**, 861–867 (2009).
41. Huang, J. *et al.* Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum. Genomics* **1**, 287–299 (2004).
42. Wang, B. *et al.* Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science* **316**, 1194–1198 (2007).
43. Schouten, J. P. *et al.* Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30**, e57 (2002).
44. Bunyan, D. J. *et al.* Dosage analysis of cancer predisposition genes by multiplex ligation-dependent probe amplification. *Br. J. Cancer* **91**, 1155–1159 (2004).
45. Fernández, L. *et al.* Comparative study of three diagnostic approaches (FISH, STRs and MLPA) in 30 patients with 22q11.2 deletion syndrome. *Clin. Genet.* **68**, 373–378 (2005).
46. Slater, H. R. *et al.* Rapid, high throughput prenatal detection of aneuploidy using a novel quantitative method (MLPA). *J. Med. Genet.* **40**, 907–912 (2003).
47. Mei, Z. *et al.* Validity of body mass index compared with other body-composition screening indexes for the assessment of body fatness in children and adolescents. *Am. J. Clin. Nutr.* **75**, 978–985 (2002).
48. Physical status: the use and interpretation of anthropometry. Report of a WHO expert committee. *World Health Organ. Tech. Rep. Ser.* **854**, 1–452 (1995).
49. Kuczmarski, R. J. *et al.* CDC growth charts: United States. *Adv. Data* **314**, 1–27 (2000).
50. Sempé, M., Pedron, G. & Roy-Pernot, M. P. *Auxologie, Méthode et Séquences*. (Thérapiex, 1979).
51. Rolland-Cachera, M. F. *et al.* Body mass index variations: centiles from birth to 87 years. *Eur. J. Clin. Nutr.* **45**, 13–21 (1991).
52. Prader, A., Largo, R. H., Molinari, L. & Issler, C. Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. *Helv. Paediatr. Acta., Suppl.* **52**, 1–125 (1989).
53. Fredriks, M. *Growth Diagrams, 1997. Fourth Dutch Nation-wide Survey.* 233–242 (Bohn Stafleu Van Loghum, 1997).
54. de Onis, M. *et al.* Development of a WHO growth reference for school-aged children and adolescents. *Bull. World Health Organ.* **85**, 660–667 (2007).
55. de Onis, M., Garza, C., Onyango, A. W. & Borghi, E. Comparison of the WHO child growth standards and the CDC 2000 growth charts. *J. Nutr.* **137**, 144–148 (2007).
56. Mei, Z., Ogden, C. L., Flegal, K. M. & Grummer-Strawn, L. M. Comparison of the prevalence of shortness, underweight, and overweight among US children aged 0 to 59 months by using the CDC 2000 and the WHO 2006 growth charts. *J. Pediatr.* **153**, 622–628 (2008).
57. Molina, J. *et al.* Abnormal social behaviors and altered gene expression rates in a mouse model for Potocki–Lupski syndrome. *Hum. Mol. Genet.* **17**, 2486–2495 (2008).
58. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{−(Delta Delta C(T))} method. *Methods* **25**, 402–408 (2001).
59. Helleman, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**, R19 (2007).
60. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 1–11 (2002).
61. Morris, J. A. & Gardner, M. J. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *Br. Med. J. (Clin. Res. Ed.)* **296**, 1313–1316 (1988).

Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk

The International Consortium for Blood Pressure Genome-Wide Association Studies

Blood pressure is a heritable trait¹ influenced by several biological pathways and responsive to environmental stimuli. Over one billion people worldwide have hypertension (≥ 140 mm Hg systolic blood pressure or ≥ 90 mm Hg diastolic blood pressure)². Even small increments in blood pressure are associated with an increased risk of cardiovascular events³. This genome-wide association study of systolic and diastolic blood pressure, which used a multi-stage design in 200,000 individuals of European descent, identified sixteen novel loci: six of these loci contain genes previously known or suspected to regulate blood pressure (*GUCY1A3–GUCY1B3*, *NPR3–C5orf23*, *ADM*, *FURIN–FES*, *GOSR2*, *GNAS–EDN3*); the other ten provide new clues to blood pressure physiology. A genetic risk score based on 29 genome-wide significant variants was associated with hypertension, left ventricular wall thickness, stroke and coronary artery disease, but not kidney disease or kidney function. We also observed associations with blood pressure in East Asian, South Asian and African ancestry individuals. Our findings provide new insights into the genetics and biology of blood pressure, and suggest potential novel therapeutic pathways for cardiovascular disease prevention.

Genetic approaches have advanced the understanding of biological pathways underlying inter-individual variation in blood pressure. For example, studies of rare Mendelian blood pressure disorders have identified multiple defects in renal sodium handling pathways⁴. More recently two genome-wide association studies (GWAS), each of >25,000 individuals of European ancestry, identified 13 loci associated with systolic blood pressure (SBP), diastolic blood pressure (DBP) and hypertension^{5,6}. We now report results of a new meta-analysis of GWAS data that includes staged follow-up genotyping to identify additional blood pressure loci.

Primary analyses evaluated associations between 2.5 million genotyped or imputed single nucleotide polymorphisms (SNPs) and SBP and DBP in 69,395 individuals of European ancestry from 29 studies (Supplementary Materials sections 1–3 and Supplementary Tables 1 and 2). Following GWAS meta-analysis, we conducted a three-stage validation experiment that made efficient use of available genotyping resources, to follow up top signals in up to 133,661 additional individuals of European descent (Supplementary Fig. 1 and Supplementary Materials section 4). Twenty-nine independent SNPs at 28 loci were significantly associated with SBP, DBP, or both in the meta-analysis combining discovery and follow-up data (Fig. 1, Table 1, Supplementary Figs 2, 3 and Supplementary Tables 3–5). All 29 SNPs attained association $P < 5 \times 10^{-9}$, an order of magnitude beyond the standard genome-wide significance level for a single-stage experiment (Table 1).

Sixteen of these 29 associations were novel (Table 1). Two associations were near the *FURIN* and *GOSR2* genes; prior targeted analyses of variants in these genes suggested they may be blood pressure loci^{7,8}. At the *CACNB2* locus we validated association for a previously reported⁶ SNP, rs4373814, and detected a novel independent association for rs1813353 (pairwise $r^2 = 0.015$ in HapMap CEU). Of our 13 previously reported associations^{5,6}, only the association at *PLCD3*

was not supported by the current results (Supplementary Table 4). Some of the associations are in or near genes involved in pathways known to influence blood pressure (*NPR3*, *GUCY1A3–GUCY1B3*, *ADM*, *GNAS–EDN3*, *NPPA–NPPB* and *CYP17A1*; Supplementary Fig. 4). Twenty-two of the 28 loci did not contain genes that were a priori strong biological candidates.

As expected from prior blood pressure GWAS results, the effects of the novel variants on SBP and DBP were small (Fig. 1 and Table 1). For all variants, the observed directions of effects were concordant for SBP, DBP and hypertension (Fig. 1, Table 1 and Supplementary Fig. 3). Among the genes at the genome-wide significant loci, only *CYP17A1*, previously implicated in Mendelian congenital adrenal hyperplasia and hypertension, is known to harbour rare variants that have large effects on blood pressure⁹.

We performed several analyses to identify potential causal alleles and mechanisms. First, we looked up the 29 genome-wide significant index SNPs and their close proxies ($r^2 > 0.8$) among *cis*-acting expression SNP (eSNP) results from multiple tissues (Supplementary Materials section 5). For 13/29 index SNPs, we found an association between nearby eSNP variants and the expression levels of at least one gene transcript ($10^{-4} > P > 10^{-51}$; Supplementary Table 6). In five cases, the index blood pressure SNP and the best eSNP from a genome-wide survey were identical, highlighting potential mediators of the SNP–blood pressure associations.

Second, because changes in protein sequence are a priori strong functional candidates, we sought non-synonymous coding SNPs that were in high linkage disequilibrium ($r^2 > 0.8$) with the 29 index SNPs. We identified such SNPs at eight loci (Table 1, Supplementary Materials section 6 and Supplementary Table 7). In addition we performed analyses testing for differences in genetic effect according to body mass index (BMI) or sex, and analyses of copy number variants, pathway enrichment and metabolomic data, but we did not find any statistically significant results (Supplementary Materials sections 7–9 and Supplementary Tables 8–10).

We evaluated whether the blood pressure variants we identified in individuals of European ancestry were associated with blood pressure in individuals of East Asian ($N = 29,719$), South Asian ($N = 23,977$) and African ($N = 19,775$) ancestries (Table 1 and Supplementary Tables 11–13). We found significant associations in individuals of East Asian ancestry for SNPs at nine loci and in individuals of South Asian ancestry for SNPs at six loci; some have been reported previously (Supplementary Tables 12 and 15). The lack of significant association for individual SNPs may reflect small sample sizes, differences in allele frequencies or linkage disequilibrium patterns, imprecise imputation for some ancestries using existing reference samples, or a genuinely different underlying genetic architecture. Because of limited power to detect effects of individual variants in the smaller non-European samples, we created genetic risk scores for SBP and DBP incorporating all 29 blood pressure variants weighted according to effect sizes observed in the European samples. In each non-European ancestry group, risk scores were strongly associated with SBP ($P = 1.1 \times 10^{-40}$ in East Asian, $P = 2.9 \times 10^{-13}$ in South Asian, $P = 9.8 \times 10^{-4}$ in African

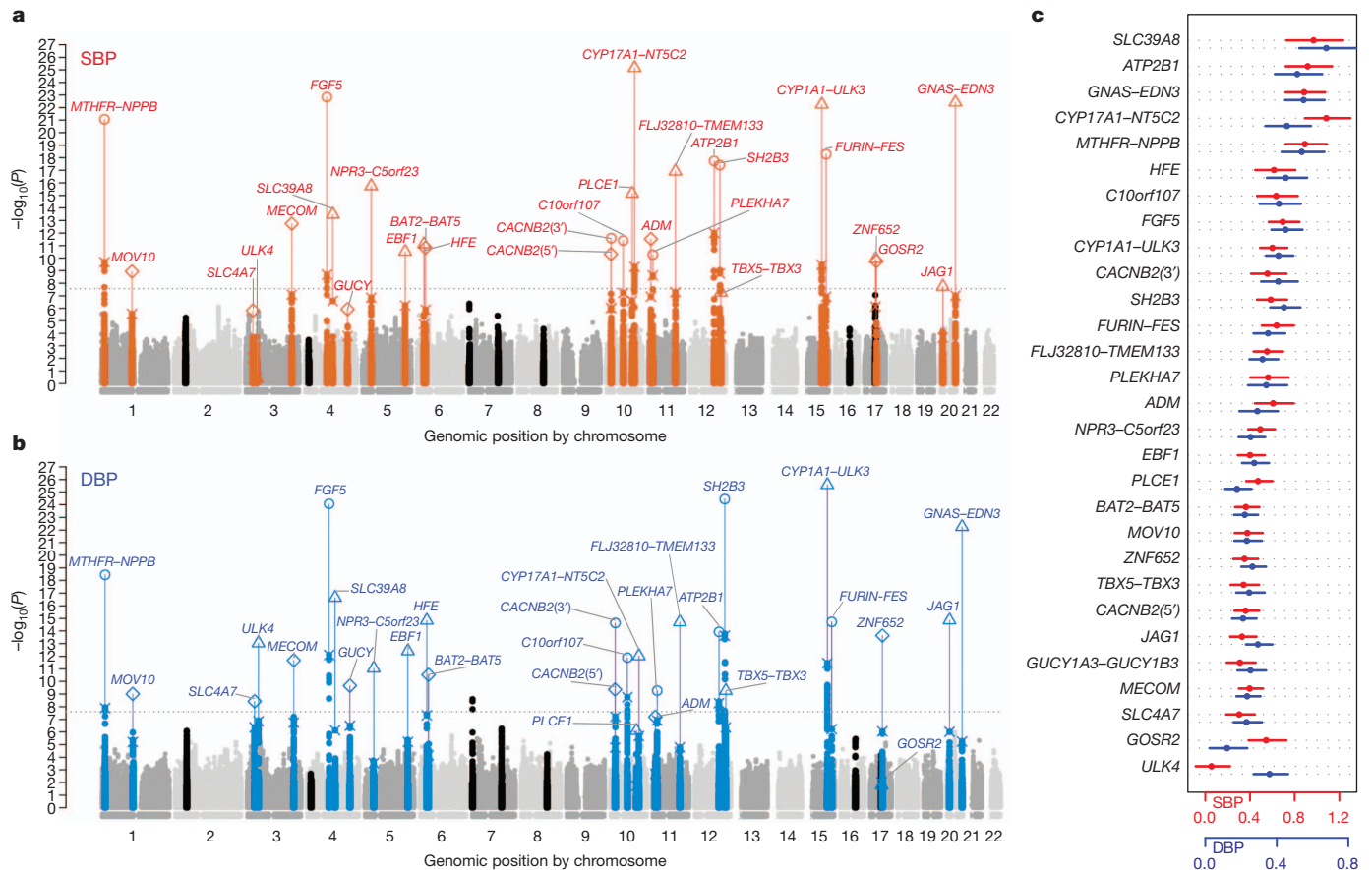


Figure 1 | Genome-wide $-\log_{10} P$ -value plots and effects for significant loci. a, b, Genome-wide $-\log_{10} P$ -value plots are shown for SBP (a) and DBP (b). SNPs within loci reaching genome-wide significance are labelled in red for SBP and blue for DBP (± 2.5 Mb of lowest P value) and lowest P values in the initial genome-wide analysis as well as the results of analysis including validation data are labelled separately. The lowest P values in the initial GWAS are denoted with a X. The range of different sample sizes in the final meta-

ancestry individuals) and DBP ($P = 2.9 \times 10^{-48}$, $P = 9.5 \times 10^{-15}$ and $P = 5.3 \times 10^{-5}$, respectively; Supplementary Table 13).

We also created a genetic risk score to assess association of the variants in aggregate with hypertension and with clinical measures of hypertensive complications including left ventricular mass, left ventricular wall thickness, incident heart failure, incident and prevalent stroke, prevalent coronary artery disease (CAD), kidney disease and measures of kidney function, using results from other GWAS consortia (Table 2, Supplementary Materials sections 10, 11 and Supplementary Table 14). The risk score was weighted using the average of SBP and DBP effects for the 29 SNPs. In an independent sample of 23,294 women¹⁰, an increase of one standard deviation in the genetic risk score was associated with a 23% increase in the odds of hypertension (95% confidence interval 19–28%; Table 2 and Supplementary Table 14). Among individuals in the top decile of the risk score, the prevalence of hypertension was 29% compared with 16% in the bottom decile (odds ratio 2.09, 95% confidence interval 1.86–2.36). Similar results were observed in an independent hypertension case-control sample (Table 2). In our study, individuals in the top compared to bottom quintiles of genetic risk score differed by 4.6 mm Hg SBP and 3.0 mm Hg DBP, differences that approach population-averaged blood pressure treatment effects for a single antihypertensive agent¹¹. Epidemiological data have shown that differences in SBP and DBP of this magnitude, across the population range of blood pressure, are associated with an increase in cardiovascular disease risk³. Consistent with this and in line with findings from randomized trials

analysis including the validation data are indicated as: circle (96,000–140,000), triangle ($>140,000$ –180,000) and diamond ($>180,000$ –220,000). SNPs near unconfirmed loci are in black. The horizontal dotted line is $P = 2.5 \times 10^{-8}$. *GUCY* denotes *GUCY1A3*–*GUCY1B3*. c, Effect size estimates and 95% confidence bars per blood-pressure-increasing allele of the 29 significant variants for SBP (red) and DBP (blue). Effect sizes are expressed in mm Hg per allele.

of blood-pressure-lowering medication in hypertensive patients^{12,13}, the genetic risk score was positively associated with left ventricular wall thickness ($P = 6.0 \times 10^{-6}$), occurrence of stroke ($P = 3.3 \times 10^{-5}$) and CAD ($P = 8.1 \times 10^{-29}$). The same genetic risk score was not, however, significantly associated with chronic kidney disease or measures of kidney function, even though these renal outcomes were available in a similar sample size as for the other outcomes (Table 2). The absence of association with kidney phenotypes could be explained by a weaker causal relationship between blood pressure and kidney phenotypes than with CAD and stroke. This finding is consistent with the mismatch between observational data that show a positive association of blood pressure with kidney disease, and clinical trial data that show inconsistent evidence of a benefit from blood pressure lowering on kidney disease prevention in patients with hypertension¹⁴. Thus, several lines of evidence converge to indicate that blood pressure elevation may in part be a consequence rather than a cause of sub-clinical kidney disease.

Our discovery meta-analysis (Supplementary Fig. 2) suggests an excess of modestly significant ($10^{-5} < P < 10^{-2}$) associations probably arising from common blood pressure variants of small effect. By dividing our principal GWAS data set into non-overlapping discovery ($N \approx 56,000$) and validation ($N \approx 14,000$) subsets, we found robust evidence for the existence of such undetected common variants (Supplementary Fig. 5 and Supplementary Materials section 12). We estimate¹⁵ that there are 116 (95% confidence interval 57–174) independent blood pressure variants with effect sizes similar to those

Table 1 | Summary association results for 29 blood pressure SNPs

Locus	Index SNP	Chr	Position	CA/ NCA	CAF	nsSNP	eSNP	SBP			DBP			HTN	
								Beta	P value	Effect in EA/SA/A	Beta	P value	Effect in EA/SA/A	Beta	P value
<i>MOV10</i>	rs2932538	1	113,018,066	G/A	0.75	Y(p)	Y(p)	0.388	1.2×10^{-9}	+/+/-	0.240	9.9×10^{-10}	+/+*/-	0.049	2.9×10^{-7}
<i>SLC4A7</i>	rs13082711	3	27,512,913	T/C	0.78	Y(p)	Y(p)	-0.315	1.5×10^{-6}	-/-/+	-0.238	3.8×10^{-9}	-/-/+	-0.035	3.6×10^{-4}
<i>MECOM</i>	rs419076	3	170,583,580	T/C	0.47	-	-	0.409	1.8×10^{-13}	+/+/+	0.241	2.1×10^{-12}	+/+/-	0.031	3.1×10^{-4}
<i>SLC39A8</i>	rs13107325	4	103,407,732	T/C	0.05	Y	Y(+)	-0.981	3.3×10^{-14}	?/+/?	-0.684	2.3×10^{-17}	?/+/?	-0.105	4.9×10^{-7}
<i>GUCY1A3- GUCY1B3</i>	rs13139571	4	156,864,963	C/A	0.76	-	-	0.321	1.2×10^{-6}	+/-/+	0.260	2.2×10^{-10}	+/-/+	0.042	2.5×10^{-5}
<i>NPR3- C5orf23</i>	rs1173771	5	32,850,785	G/A	0.60	-	-	0.504	1.8×10^{-16}	+*/+/+	0.261	9.1×10^{-12}	+*/+/-	0.062	3.2×10^{-10}
<i>EBF1</i>	rs11953630	5	157,777,980	T/C	0.37	-	-	-0.412	3.0×10^{-11}	+/+/+	-0.281	3.8×10^{-13}	+/+/+	-0.052	1.7×10^{-7}
<i>HFE</i>	rs1799945	6	26,199,158	G/C	0.14	Y	-	0.627	7.7×10^{-12}	+/+/-	0.457	1.5×10^{-15}	+/+/-	0.095	1.8×10^{-10}
<i>BAT2-BAT5</i>	rs805303	6	31,724,345	G/A	0.61	Y(p)	Y(+)	0.376	1.5×10^{-11}	-/-/?	0.228	3.0×10^{-11}	-/-/+	0.054	1.1×10^{-10}
<i>CACNB2(5')</i>	rs4373814	10	18,459,978	G/C	0.55	-	-	-0.373	4.8×10^{-11}	+/+/-	-0.218	4.4×10^{-10}	+/-/+	-0.046	8.5×10^{-8}
<i>PLCE1</i>	rs932764	10	95,885,930	G/A	0.44	-	-	0.484	7.1×10^{-16}	+/+/-	0.185	8.1×10^{-7}	+/+/-	0.055	9.4×10^{-9}
<i>ADM</i>	rs7129220	11	10,307,114	G/A	0.89	-	-	-0.619	3.0×10^{-12}	?/-/+	-0.299	6.4×10^{-8}	?/-/+	-0.044	1.1×10^{-3}
<i>FLJ32810- TMEM133</i>	rs633185	11	100,098,748	G/C	0.28	-	-	-0.565	1.2×10^{-17}	+*/+/+	-0.328	2.0×10^{-15}	+*/+/-	-0.070	5.4×10^{-11}
<i>FURIN-FES</i>	rs2521501	15	89,238,392	T/A	0.31	-	Y(-)	0.650	5.2×10^{-19}	+*/+/+	0.359	1.9×10^{-15}	+*/+/+	0.059	7.0×10^{-7}
<i>GOSR2</i>	rs17608766	17	42,368,270	T/C	0.86	-	Y(+)	-0.556	1.1×10^{-10}	+/-/+	-0.129	0.017	+/-/+	-0.025	0.08
<i>JAG1</i>	rs1327235	20	10,917,030	G/A	0.46	-	-	0.340	1.9×10^{-8}	+*/+/+	0.302	1.4×10^{-15}	+*/+*/+	0.034	4.6×10^{-4}
<i>GNAS-EDN3</i>	rs6015450	20	57,184,512	G/A	0.12	Y(p)	-	0.896	3.9×10^{-23}	?/+/?	0.557	5.6×10^{-23}	?/+*/+	0.110	4.2×10^{-14}
<i>MTHFR- NPPB</i>	rs17367504	1	11,785,365	G/A	0.15	-	Y(-/r)	-0.903	8.7×10^{-22}	+/+/+	-0.547	3.5×10^{-19}	+/+/+	-0.103	2.3×10^{-10}
<i>ULK4</i>	rs3774372	3	41,852,418	T/C	0.83	Y	Y(r/p)	-0.067	0.39	-/-/+	-0.367	9.0×10^{-14}	+/+/+	-0.017	0.18
<i>FGF5</i>	rs1458038	4	81,383,747	T/C	0.29	-	-	0.706	1.5×10^{-23}	+*/+/+	0.457	8.5×10^{-25}	+*/+*/+	0.072	1.9×10^{-7}
<i>CACNB2(3')</i>	rs1813353	10	18,747,454	T/C	0.68	-	-	0.569	2.6×10^{-12}	+/+/+	0.415	2.3×10^{-15}	+/+/+	0.078	6.2×10^{-10}
<i>C10orf107</i>	rs4590817	10	63,137,559	G/C	0.84	-	Y(r)	0.646	4.0×10^{-12}	-/+/-	0.419	1.3×10^{-12}	-/-/-	0.096	9.8×10^{-9}
<i>CYP17A1- NT5C2</i>	rs11191548	10	104,836,168	T/C	0.91	-	Y(-)	1.095	6.9×10^{-26}	+*/+*/+	0.464	9.4×10^{-13}	+*/+*/+	0.097	1.4×10^{-5}
<i>PLEKHA7</i>	rs381815	11	16,858,844	T/C	0.26	-	-	0.575	5.3×10^{-11}	+*/+/+	0.348	5.3×10^{-10}	+*/-/+	0.062	3.4×10^{-6}
<i>ATP2B1</i>	rs17249754	12	88,584,717	G/A	0.84	-	-	0.928	1.8×10^{-18}	+*/+*/-	0.522	1.2×10^{-14}	+*/+*/-	0.126	1.1×10^{-14}
<i>SH2B3</i>	rs3184504	12	110,368,991	T/C	0.47	Y	Y(+)	0.598	3.8×10^{-18}	-/-/+	0.448	3.6×10^{-25}	-/-/+	0.056	2.6×10^{-6}
<i>TBX5-TBX3</i>	rs10850411	12	113,872,179	T/C	0.7	-	-	0.354	5.4×10^{-8}	-/+/-	0.253	5.4×10^{-10}	-/-/-	0.045	5.2×10^{-6}
<i>CYP1A1- ULK3</i>	rs1378942	15	72,864,420	C/A	0.35	-	Y(+)	0.613	5.7×10^{-23}	+*/+/+	0.416	2.7×10^{-26}	+*/+/-	0.073	1.0×10^{-8}
<i>ZNF652</i>	rs12940887	17	44,757,806	T/C	0.38	-	Y(-)	0.362	1.8×10^{-10}	+/-/+	0.27	2.3×10^{-14}	+/-/+	0.046	1.2×10^{-7}

Summary association statistics, based on combined discovery and follow-up data, for 29 independent SNPs in individuals of European ancestry are shown. New genome-wide significant findings (17 SNPs) are presented in the top half of the table, data on 12 previously published signals are presented in the lower half. Y indicates that the blood pressure index SNP is a non-synonymous (ns)SNP, Y(p) indicates a proxy SNP is a nsSNP. Y(+) indicates that the blood pressure index SNP is the strongest known eSNP for a transcript; Y(-) indicates that the blood pressure index SNP is an eSNP but not the strongest known eSNP for any transcript. Y(r) indicates that the blood pressure index SNP is the strongest known eSNP in a targeted real-time PCR experiment. Y(p) indicates that a proxy SNP ($r^2 > 0.8$) to a blood pressure SNP is an eSNP but not the strongest known eSNP. Observed effect directions in East Asian (EA), South Asian (SA) and African (A) ancestry individuals are coded + or - if concordant or discordant with directions in European ancestry results. Effect size estimates (beta) correspond to mm Hg per coded allele for SBP and DBP and ln(odds) per coded allele for hypertension (HTN). CA, coded allele; CAF, coded allele frequency; NCA, non-coded allele. ? denotes missing data. Genomic positions use NCBI Build 36 coordinates.

* Significant, controlling the FDR at 5% over 58 tests per ancestry (Supplementary Tables 5 and 12).

reported here, which collectively can explain ~2.2% of the phenotypic variance for SBP and DBP, compared with 0.9% explained by the 29 associations discovered thus far (Supplementary Fig. 6 and Supplementary Materials section 13).

Most of the 28 blood pressure loci harbour multiple genes (Supplementary Table 15 and Supplementary Fig. 4), and although substantial research is required to identify the specific genes and variants responsible for these associations, several loci contain highly plausible biological candidates. The *NPPA* and *NPPB* genes at the *MTHFR-NPPB* locus encode precursors for atrial- and B-type natriuretic peptides (ANP, BNP), and previous work has identified SNPs—modestly correlated with our index SNP at this locus—which are associated with plasma ANP, BNP and blood pressure¹⁶. We found the index SNP at this locus was associated with opposite effects on blood pressure and on ANP/BNP levels, consistent with a model in which the variants act through increased ANP/BNP production to lower blood pressure¹⁶ (Supplementary Materials section 14).

Two other loci identified in the current study harbour genes involved in natriuretic peptide and related nitric oxide signalling pathways^{17,18}, both of which act to regulate cyclic guanosine monophosphate. The first locus contains *NPR3*, which encodes the natriuretic peptide clearance receptor (NPR-C). *NPR3* knockout mice exhibit reduced clearance of circulating natriuretic peptides and lower blood pressure¹⁹. The second locus includes *GUCY1A3* and *GUCY1B3*, encoding the α and β subunits of soluble guanylate cyclase; knockout of either gene in murine models results in hypertension²⁰.

Another locus contains *ADM*—encoding adrenomedullin—which has natriuretic, vasodilatory and blood-pressure-lowering properties²¹. At the *GNAS-EDN3* locus, *ZNF831* is closest to the index SNP, but *GNAS* and *EDN3* are two nearby compelling biological candidates (Supplementary Fig. 4 and Supplementary Table 15).

We identified two loci with plausible connections to blood pressure via genes implicated in renal physiology or kidney disease. At the first locus, *SLC4A7* is an electro-neutral sodium bicarbonate co-transporter expressed in the nephron and in vascular smooth muscle²². At the second locus, *PLCE1* (phospholipase-C-epsilon-1 isoform) is important for normal podocyte development in the glomerulus; sequence variation in *PLCE1* has been implicated in familial nephrotic syndromes and end-stage kidney disease²³.

Missense variants in two genes involved in metal ion transport were associated with blood pressure in our study. The first encodes a His/Asp change at amino acid 63 (H63D) in *HFE* and is a low-penetrance allele for hereditary hemochromatosis²⁴. The second is an Ala/Thr polymorphism located in exon 7 of *SLC39A8*, which encodes a zinc transporter that also transports cadmium and manganese²⁵. The same allele of *SLC39A8* associated with blood pressure in our study has recently been associated with high-density lipoprotein cholesterol levels²⁶ and BMI²⁷ (Supplementary Table 15).

We have shown that 29 independent genetic variants influence blood pressure in people of European ancestry. The variants reside in 28 loci, 16 of which were novel, and we confirmed association of several of them in individuals of non-European ancestry. A risk score

Table 2 | Genetic risk score and cardiovascular outcome association results

Phenotype	Source	Effect		s.e.	P value	No. SNPs	Contrast top versus bottom			N case/control or total
		(per s.d. of genetic risk score)					Quintiles	Deciles		
Blood pressure phenotypes										
SBP (mm Hg)	WGHS	1.645	0.098	(a)	6.5×10^{-63}	29	4.61	5.77	(a)	23,294
DBP (mm Hg)	WGHS	1.057	0.067	(a)	8.4×10^{-57}	29	2.96	3.71	(a)	23,294
Prevalent hypertension	WGHS	0.211	0.018	(b)	3.1×10^{-33}	29	1.80	2.09	(b)	5,018/18,276
Prevalent hypertension	BRIGHT	0.287	0.031	(b)	7.7×10^{-21}	29	2.23	2.74	(b)	2,406/1,990
Dichotomous endpoints										
Incident heart failure	CHARGE-HF	0.035	0.021	(c)	0.10	29	1.10	1.13	(c)	2,526/18,400
Incident stroke	NEURO-CHARGE	0.103	0.028	(c)	0.0002	28	1.34	1.44	(c)	1,544/18,058
Prevalent stroke	SCG	0.075	0.037	(b)	0.05	29	1.23	1.30	(b)	1,473/1,482
Stroke (combined, incident and prevalent)	CHARGE & SCG	NA	NA	NA	3.3×10^{-5}	NA	NA	NA	NA	3,017/19,540
Prevalent CAD	CARDIoGRAM	0.092	0.010	(b)	1.6×10^{-19}	28	1.29	1.38	(b)	22,233/64,726
Prevalent CAD	C4D ProCARDIS	0.132	0.022	(b)	2.2×10^{-9}	29	1.45	1.59	(b)	5,720/4,381
Prevalent CAD	C4D HPS	0.083	0.027	(b)	0.002	29	1.26	1.34	(b)	2,704/2,804
Prevalent CAD (combined)	CARDIoGRAM & C4D	0.100	0.009	(b)	8.1×10^{-29}	29	1.32	1.42	(b)	30,657/71,911
Prevalent chronic kidney disease	CKDGen	0.014	0.015	(b)	0.35	29	1.04	1.05	(b)	5,807/61,286
Prevalent microalbuminuria	CKDGen	0.008	0.019	(b)	0.68	29	1.02	1.03	(b)	3,698/27,882
Continuous measures of target organ damage										
Left ventricular mass (g)	EchoGen	0.822	0.317	(a)	0.01	29	2.30	2.89	(a)	12,612
Left ventricular wall thickness (cm)	EchoGen	0.009	0.002	(a)	6.0×10^{-6}	29	0.03	0.03	(a)	12,612
Serum creatinine	KidneyGen	-0.001	0.001	(d)	0.24	29	1.00	1.00	(d)	23,812
eGFR (four-parameter MDRD equation)	CKDGen	-0.0001	0.0009	(d)	0.93	29	1.00	1.00	(d)	67,093
Urinary albumin/creatinine ratio	CKDGen	0.005	0.007	(d)	0.43	29	1.01	1.02	(d)	31,580

Association of genetic risk score (using all 29 SNPs at 28 loci, parameterized using the average of SBP and DBP effects (= (SBP effect + DBP effect)/2) from the discovery analysis), tested in results from other GWAS consortia. (a) Units are the unit of phenotypic measurement, either per standard deviation (s.d.) of genetic risk score, or as a difference between top/bottom quintiles or deciles. (b) Units are ln(odds) per s.d. of genetic risk score, or odds ratio between top/bottom quintiles or deciles. (c) Units are ln(hazard) per s.d. of genetic risk score, or hazard ratio between top/bottom quintiles or deciles. (d) Units are ln(phenotype) per s.d. of genetic risk score, or phenotypic ratio between top/bottom quintiles or deciles. s.e., standard error. SCG, UK-US Stroke Collaborative Group; see Supplementary Materials sections 1.79 and 11 for further detail on consortia and studies.

derived from the 29 variants was significantly associated with blood-pressure-related organ damage and clinical cardiovascular disease, but not kidney disease. These loci improve our understanding of the genetic architecture of blood pressure, provide new biological insights into blood pressure control and may identify novel targets for the treatment of hypertension and the prevention of cardiovascular disease.

Note added in proof: Since this manuscript was submitted, Kato *et al.* published a blood pressure GWAS in East Asians that identified a SNP highly correlated to the SNP we report at the *NPR3/C5orf23* locus²⁸.

METHODS SUMMARY

Supplementary Materials provide complete methods and include the following sections: study recruitment and phenotyping, adjustment for antihypertensive medications, genotyping, data quality control, genotype imputation, within-cohort association analyses, meta-analyses of discovery and validation stages, stratified analyses by sex and BMI, identification of eSNPs and non-synonymous SNPs, metabolomic and lipidomic analyses, CNV analyses, pathway analyses, analyses for non-European ancestries, association of a risk score with hypertension and cardiovascular disease, estimation of numbers of undiscovered variants, measurement of natriuretic peptides, and brief literature reviews and GWAS database lookups of all validated blood pressure loci. Full GWAS results for ≈ 2.5 million SNPs are also provided.

Received 16 August 2010; accepted 28 July 2011.

Published online 11 September 2011.

- Levy, D. *et al.* Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the Framingham heart study. *Hypertension* **36**, 477–483 (2000).
- Kearney, P. M. *et al.* Global burden of hypertension: analysis of worldwide data. *Lancet* **365**, 217–223 (2005).
- Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* **360**, 1903–1913 (2002).
- Lifton, R. P., Gharavi, A. G. & Geller, D. S. Molecular mechanisms of human hypertension. *Cell* **104**, 545–556 (2001).
- Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genet.* **41**, 666–676 (2009).
- Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nature Genet.* **41**, 677–687 (2009).

- Meyer, T. E. *et al.* GOSR2 Lys67Arg is associated with hypertension in whites. *Am. J. Hypertens.* **22**, 163–168 (2009).
- Li, N. *et al.* Associations between genetic variations in the *FURIN* gene and hypertension. *BMC Med. Genet.* **11**, 124 (2010).
- Mussig, K. *et al.* 17 α -hydroxylase/17,20-lyase deficiency caused by a novel homozygous mutation (Y27Stop) in the cytochrome CYP17 gene. *J. Clin. Endocrinol. Metab.* **90**, 4362–4365 (2005).
- Ridker, P. M. *et al.* Rationale, design, and methodology of the Women's Genome Health Study: a genome-wide association study of more than 25,000 initially healthy american women. *Clin. Chem.* **54**, 249–255 (2008).
- Burt, V. L. *et al.* Trends in the prevalence, awareness, treatment, and control of hypertension in the adult US population. Data from the health examination surveys, 1960 to 1991. *Hypertension* **26**, 60–69 (1995).
- Blood Pressure Lowering Treatment Trialists' Collaboration Effects of different regimens to lower blood pressure on major cardiovascular events in older and younger adults: meta-analysis of randomised trials. *Br. Med. J.* **336**, 1121–1123 (2008).
- Law, M. R., Morris, J. K. & Wald, N. J. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. *Br. Med. J.* **338**, b1665 (2009).
- Lewis, J. B. Blood pressure control in chronic kidney disease: is less really more? *J. Am. Soc. Nephrol.* **21**, 1086–1092 (2010).
- Park, J. H. *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature Genet.* **42**, 570–575 (2010).
- Newton-Cheh, C. *et al.* Association of common variants in *NPPA* and *NPPB* with circulating natriuretic peptides and blood pressure. *Nature Genet.* **41**, 348–353 (2009).
- Schenk, D. B. *et al.* Purification and subunit composition of atrial natriuretic peptide receptor. *Proc. Natl Acad. Sci. USA* **84**, 1521–1525 (1987).
- Schmidt, H. H. & Walter, U. NO at work. *Cell* **78**, 919–925 (1994).
- Matsukawa, N. *et al.* The natriuretic peptide clearance receptor locally modulates the physiological effects of the natriuretic peptide system. *Proc. Natl Acad. Sci. USA* **96**, 7403–7408 (1999).
- Friebe, A., Mergia, E., Dangel, O., Lange, A. & Koesling, D. Fatal gastrointestinal obstruction and hypertension in mice lacking nitric oxide-sensitive guanylyl cyclase. *Proc. Natl Acad. Sci. USA* **104**, 7699–7704 (2007).
- Ishimitsu, T., Ono, H., Minami, J. & Matsuoka, H. Pathophysiologic and therapeutic implications of adrenomedullin in cardiovascular disorders. *Pharmacol. Ther.* **111**, 909–927 (2006).
- Pushkin, A. *et al.* Cloning, tissue distribution, genomic organization, and functional characterization of NBC3, a new member of the sodium bicarbonate cotransporter family. *J. Biol. Chem.* **274**, 16569–16575 (1999).
- Hinkes, B. *et al.* Positional cloning uncovers mutations in *PLCE1* responsible for a nephrotic syndrome variant that may be reversible. *Nature Genet.* **38**, 1397–1405 (2006).

24. Feder, J. N. *et al.* A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nature Genet.* **13**, 399–408 (1996).
25. He, L., Wang, B., Hay, E. B. & Nebert, D. W. Discovery of ZIP transporters that participate in cadmium damage to testis and kidney. *Toxicol. Appl. Pharmacol.* **238**, 250–257 (2009).
26. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
27. Spelioti, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42**, 937–948 (2010).
28. Kato, N. *et al.* Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nature Genet.* **43**, 531–538 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A number of the participating studies and authors are members of the CHARGE and Global BPgen consortia. Many funding mechanisms by NIH/NHLBI, European and private funding agencies contributed to this work and a full list is provided in section 21 of the Supplementary Materials.

Author Contributions Full author contributions and roles are listed in Supplementary Materials section 19.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.C. (aravinda@jhmi.edu), M.C. (m.j.caulfield@qmul.ac.uk), D.L. (levy@nhlbi.nih.gov), P.B.M. (p.b.munroe@qmul.ac.uk), C.N.-C. (cnewtonch@chgr.mgh.harvard.edu).

George B. Ehret^{1,2,3*}, Patricia B. Munroe^{4*}, Kenneth M. Rice^{5*}, Murielle Bochud^{2*}, Andrew D. Johnson^{6,7*}, Daniel I. Chasman^{8,9*}, Albert V. Smith^{10,11*}, Martin D. Tobin¹², Germaine C. Verwoert^{13,14,15}, Shih-Jen Hwang^{6,7,16}, Vasyli Pihur¹, Peter Vollenweider¹⁷, Paul F. O'Reilly¹⁸, Najaf Amin¹³, Jennifer L. Bragg-Gresham¹⁹, Alexander Teumer²⁰, Nicole L. Glazer²¹, Lenore Launer²², Jing Hua Zhao²³, Yuri Aulchenko¹³, Simon Heath²⁴, Sigmund Söber²⁵, Afshin Parsa²⁶, Jian'an Luan²³, Pankaj Arora²⁷, Abbas Dehghan^{13,14,15}, Feng Zhang²⁸, Gavin Lucas²⁹, Andrew A. Hicks³⁰, Anne U. Jackson³¹, John F. Peden³², Toshiko Tanaka³³, Sarah H. Wild³⁴, Igor Rudan^{35,36}, Wilmar Igl³⁷, Yuri Milaneschi³⁸, Alex N. Parker³⁸, Cristiano Fava^{39,40}, John C. Chambers^{18,41}, Ervin R. Fox⁴², Meena Kumari⁴³, Min Jin Go⁴⁴, Pim van der Harst⁴⁵, Wen Hong Linda Kao⁴⁶, Marketa Sjögren³⁹, D. G. Vinay⁴⁷, Myriam Alexander⁴⁸, Yasuharu Tabara⁴⁹, Sue Shaw-Hawkins⁵⁰, Peter H. Whincup⁵⁰, Yongmei Liu⁵¹, Gang Shi⁵², Johanna Kuusisto⁵³, Bamidele Tayo⁵⁴, Mark Seielstad^{55,56}, Xuelling Sim⁵⁷, Khanh-Dung Hoang Nguyen¹, Terho Lehtimäki⁵⁸, Giuseppe Matullo^{59,60}, Ying Wu⁶¹, Tom R. Gaunt⁶², N. Charlotte Omland-Orsted^{63,64}, Matthew N. Cooper⁶⁵, Carl G. P. Platou⁶⁶, Elin Org⁶⁵, Rebecca Hardy⁶⁷, Santosh Dahgam⁶⁸, Jutta Palmer⁶⁹, Veronique Vitart⁷⁰, Peter S. Braund^{71,72}, Tatiana Kuznetsova⁷³, Cuno S. P. M. Uiterwaal⁶³, Adebawale Adeyemo⁷⁴, Walter Palmas⁷⁵, Harry Campbell³⁵, Barbara Ludwig⁷⁶, Maciej Tomaszewski^{77,78}, Ioanna Tzoulaki^{77,78}, Nicholette D. Palmer⁷⁹, CARDIoGRAM consortium⁸⁰, CKDGen Consortium⁸¹, KidneyGen Consortium⁸², EchoGen consortium⁸³, CHARGE-HF consortium⁸⁴, Thor Aspelund^{10,11}, Melissa Garcia²², Yen-Pei C. Chang²⁶, Jeffrey R. O'Connell²⁶, Nanette I. Steinle²⁶, Diederick E. Grobbee⁸⁵, Dan E. Arking⁸⁶, Sharon L. Kardina⁸⁰, Alanna C. Morrison⁸¹, Dena Hernandez⁸², Samer Najjar^{83,84}, Wendy L. McArdle⁸⁵, David Hadley^{50,86}, Morris J. Brown⁸⁷, John M. Connell⁸⁸, Aaron D. Hingorani⁸⁹, Ian N. M. Day⁶², Debbie A. Lawlor⁶², John P. Beilby^{90,91}, Robert W. Lawrence⁶⁵, Robert Clarke⁹², Jemma C. Hopewell⁹², Halit Ongen³², Albert W. Dreisbach⁴², Yali Li⁹³, J. Hunter Young⁹⁴, Joshua C. Bis²¹, Mika Kahönen⁹⁵, Jorma Viikari⁹⁶, Linda S. Adair⁹⁷, Nanette R. Lee⁹⁸, Ming-Huei Chen⁹⁹, Matthias Olden^{100,101}, Cristian Pattaro³⁰, Judith A. Hoffman Bolton¹⁰², Anna Köttgen^{102,103}, Sven Bergmann^{104,105}, Vincent Mooser¹⁰⁶, Nish Chaturvedi¹⁰⁷, Timothy M. Frayling¹⁰⁸, Muhammad Islam¹⁰⁹, Tazeen H. Jafar¹⁰⁹, Jeanette Erdmann¹¹⁰, Smita R. Kulkarni¹¹¹, Stefan R. Bornstein⁷⁶, Jürgen Grässler⁷⁶, Leif Groop^{112,113}, Benjamin F. Voight¹¹⁴, Johannes Kettunen^{115,116}, Philip Howard¹¹⁷, Andrew Taylor⁴³, Simonetta Guarrera⁶⁰, Fulvio Ricceri^{59,60}, Valur Emilsson¹¹⁸, Andrew Plump¹¹⁸, Inês Barroso^{119,120}, Kay-Tee Khaw⁴⁸, Alan B. Weder¹²¹, Steven C. Hunt¹²², Yan V. Sun⁸⁰, Richard N. Bergman¹²³, Francis S. Collins¹²⁴, Lori L. Bonnycastle¹²⁴, Laura J. Scott³¹, Heather M. Stringham³¹, Leena Peltonen^{116,119,125,126}, Markus Perola¹²⁵, Erkki Vartiainen¹²⁵, Stefan-Martin Brand^{127,128}, Jan A. Staessen⁷³, Thomas J. Wang^{6,129}, Paul R. Burton^{127,128}, Maria Soler Artigas¹², Yanbin Dong¹³⁰, Harold Snieder^{130,131}, Xiaoling Wang¹³⁰, Haidong Zhu¹³⁰, Kurt K. Lohman¹³², Megan E. Rudock³¹, Susan R. Heckbert^{133,134}, Nicholas L. Smith^{133,134,135}, Kerri L. Wiggins¹³⁶, Ayo Doughty⁷⁴, Daniel Shrier⁷⁴, Gudrun Veldre^{25,137}, Margus Viigimaa^{138,139}, Sanjay Kinra¹⁴⁰, Dorairaj Prabhakaran¹⁴¹, Vikal Tripathy¹⁴¹, Carl D. Langefeld⁷⁹, Annika Rosengren¹⁴², Dag S. Thelle¹⁴³, Anna Maria Corsi¹⁴⁴, Andrew Singleton⁸², Terrence Forrester¹⁴⁵, Gina Hilton¹, Colin A. McKenzie¹⁴⁵, Tunde Salako¹⁴⁶, Naoharu Iwai¹⁴⁷, Yoshikuni Kita¹⁴⁸, Toshio Ogihara¹⁴⁹, Takayoshi Ohkubo^{148,150}, Tomonori Okamura^{147,148}, Hirotsugu Ueshima^{148,151}, Satoshi Umemura¹⁵², Susana Eyheramendy¹⁵³, Thomas Meitinger^{154,155}, H.-Erich Wichmann^{156,157,158}, Yoon Shin Cho⁴⁴, Hyung-Lae Kim⁴⁴, Jong-Young Lee⁴⁴, James Scott¹⁵⁹, Joban S. Sehmi^{41,159}, Weihua Zhang¹⁶⁰, Bo Hedblad³⁹, Peter Nilsson³⁹, George Davey Smith⁶², Andrew Wong⁶⁷, Narisu Narisu¹²⁴, Alena Stančáková⁵³, Leslie J. Raffel¹⁶⁰, Jie Yao¹⁶⁰, Sekar Kathiresan^{27,161}, Christopher J. O'Donnell^{9,27,162}, Stephen M. Schwartz¹³³, M. Arfan Ikram^{13,15}, W. T. Longstreth Jr¹⁶³, Thomas H. Mosley¹⁶⁴, Sudha Seshadri¹⁶⁵, Nick R.G. Shrine¹², Louise V. Wain¹²,

Mario A. Morken¹²⁴, Amy J. Swift¹²⁴, Jaana Laitinen¹⁶⁶, Inga Prokopenko^{51,167}, Paavo Zitting¹⁶⁸, Jackie A. Cooper⁶⁹, Steve E. Humphries⁶⁹, John Danesh⁴⁸, Asif Rasheed¹⁶⁹, Anuj Goel³², Anders Hamsten¹⁷⁰, Hugh Watkins³², Stephan J. L. Bakker¹⁷¹, Wiek H. van Gilst⁴⁵, Charles S. Janipalli⁴⁷, K. Radha Mani⁴⁷, Chittaranjan S. Yajnik¹¹¹, Albert Hofman¹³, Francesco U. S. Mattace-Raso^{13,14}, Ben A. Oostra¹⁷², Ayse Demirkan¹³, Aaron Isaacs¹³, Fernando Rivadeneira^{13,14}, Edward G. Lakatta¹⁷³, Marco Orru^{174,175}, Angelo Scuteri¹⁷³, Mika Ala-Korpela^{176,177,178}, Antti J. Kangas¹⁷⁶, Leo-Pekka Lytikäinen¹⁷⁸, Pasi Soininen^{176,177}, Taru Tuikainen^{176,179,180}, Peter Würtz^{181,176,179}, Rick Twee-Hee Ong^{56,57,181}, Marcus Dörner¹⁸², Heyo K. Kroemer¹⁸³, Uwe Völker²⁰, Henry Völzke¹⁸⁴, Pilar Galan¹⁸⁵, Serge Hercberg¹⁸⁵, Mark Lathrop²⁴, Diana Zelenika²⁴, Panos Deloukas¹¹⁹, Massimo Mangino²⁸, Tim D. Spector²⁸, Guangju Zhai²⁸, James F. Meschia¹⁸⁶, Michael A. Nalls⁸², Pankaj Sharma¹⁸⁷, Janos Terzic¹⁸⁸, M. V. Kranthi Kumar⁴⁷, Matthew Denniff⁷¹, Ewa Zukowska-Szczeczowska¹⁸⁹, Lynne E. Wagenknecht⁷⁹, F. Gerald R. Fowkes¹⁹⁰, Fadi J. Charchar¹⁹¹, Peter E. H. Schwarz¹⁹², Caroline Hayward⁷⁰, Xiuqing Guo¹⁶⁰, Charles Rotimi⁷⁴, Michiel L. Bots⁶³, Eva Brand¹⁹³, Nilesh J. Samani^{17,72}, Ozren Polasek¹⁹⁴, Philippa J. Talmud⁶⁰, Fredrik Nyberg^{68,195}, Diana Kuh⁶⁷, Maris Laan²⁵, Kristian Hveem⁶⁶, Lyle J. Palmer^{196,197}, Yvonne T. van der Schouw⁶³, Juan P. Casas¹⁹⁸, Karen L. Mohlke⁶¹, Paolo Vineis^{60,199}, Olli Raitakari²⁰⁰, Santhi K. Ganesh²⁰¹, Tien Y. Wong^{202,203}, E. Shyong Tai^{57,204,205}, Richard S. Cooper⁵⁴, Markku Laakso⁵³, Dabeeru C. Rao²⁰⁶, Tamara B. Harris²², Richard W. Morris²⁰⁷, Anna F. Dominiczak²⁰⁸, Mika Kivimäki²⁰⁹, Michael G. Marmot²⁰⁹, Tetsuro Miki⁴⁹, Danish Saleheen^{48,169}, Giriraj R. Chandak⁴⁷, Josef Coresh²¹⁰, Gerjan Navis²¹¹, Veikko Salomaa¹²⁵, Bok-Ghee Han⁴⁴, Xiaofeng Zhu⁹³, Jaspal S. Kooner^{41,159}, Olle Melander³⁹, Paul M. Ridker^{8,9,212}, Stefania Bandinelli²¹³, Ulf B. Gyllenstein³⁷, Alan F. Wright⁷⁰, James F. Wilson³⁴, Luigi Ferrucci³³, Martin Farrall³², Jaakko Tuomilehto^{214,215,216,217}, Peter P. Pramstaller^{30,218}, Roberto Elosua^{29,219}, Nicole Soranzo^{28,119}, Eric J. G. Sijbrands^{13,14}, David Altshuler^{114,220}, Ruth J. F. Loos²³, Alan R. Shuldiner^{26,221}, Christian Gieger¹⁵⁶, Pierre Meneton²²², Andre G. Uitterlinden^{13,14,15}, Nicholas J. Wareham²³, Vilmondur Gudnason^{10,11}, Jerome I. Rotter¹⁶⁰, Rainer Rettig²²³, Manuela Uda¹⁷⁴, David P. Strachan⁵⁰, Jacqueline C. M. Witterman^{13,15}, Anna-Liisa Hartikainen²²⁴, Jacques S. Beckmann^{104,225}, Eric Boerwinkle²²⁶, Ramachandran S. Vasan^{6,227}, Michael Boehnke³¹, Martin G. Larson^{6,228}, Marjo-Riitta Järvelin^{18,229,230,231,232}, Bruce M. Psaty^{21,134*}, Gonçalo R. Abecasis^{19*}, Aravinda Chakravarti^{1*}, Paul Elliott^{18,232*}, Cornelia M. van Duijn^{13,233*}, Christopher Newton-Cheh^{27,114*}, Daniel Levy^{6,7,16*}, Mark J. Caulfield^{4*} & Toby Johnson^{4*}

¹Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ²Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois and University of Lausanne, Bugnion 17, 1005 Lausanne, Switzerland. ³Cardiology, Department of Specialties of Internal Medicine, Geneva University Hospital, Rue Gabrielle-Perret-Gentil 4, 1211 Geneva 14, Switzerland. ⁴Clinical Pharmacology and The Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. ⁵Department of Biostatistics, University of Washington, Seattle, Washington 98195, USA. ⁶Framingham Heart Study, Framingham, Massachusetts 01702, USA. ⁷National Heart Lung, and Blood Institute, Bethesda, Maryland 20824, USA. ⁸Division of Preventive Medicine, Brigham and Women's Hospital, 900 Commonwealth Avenue East, Boston, Massachusetts 02215, USA. ⁹Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁰Icelandic Heart Association, 201 Kópavogur, Iceland. ¹¹University of Iceland, 101 Reykjavik, Iceland. ¹²Department of Health Sciences, University of Leicester, University Rd, Leicester LE1 7RH, UK. ¹³Department of Epidemiology, Erasmus Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands. ¹⁴Department of Internal Medicine, Erasmus Medical Center, 3000 CA Rotterdam, The Netherlands. ¹⁵Netherlands Consortium for Healthy Aging (NCHA), Netherland Genome Initiative (NGI), Erasmus 3000 CA Rotterdam, The Netherlands. ¹⁶Center for Population Studies, National Heart Lung, and Blood Institute, Bethesda, Maryland 20824, USA. ¹⁷Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland. ¹⁸Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK. ¹⁹Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan 48103, USA. ²⁰Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany. ²¹Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, Washington 98101, USA. ²²Laboratory of Epidemiology, Demography, Biometry, National Institute on Aging, National Institutes of Health, Bethesda, Maryland 20892, USA. ²³MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge CB2 0QQ, UK. ²⁴Centre National de Génétique, Commissariat à l'Energie Atomique, Institut de Génétique, 91057 Evry, France. ²⁵Institute of Molecular and Cell Biology, University of Tartu, Riia 23, Tartu 51010, Estonia. ²⁶University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. ²⁷Center for Human Genetic Research, Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ²⁸Department of Twin Research & Genetic Epidemiology, King's College London, London SE1 7EH, UK. ²⁹Cardiovascular Epidemiology and Genetics, Institut Municipal d'Investigació Mèdica, Barcelona Biomedical Research Park, 88 Doctor Aiguader, 08003 Barcelona, Spain. ³⁰Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Viale Druso 1, 39100 Bolzano, Italy - Affiliated Institute of the University of Lübeck, Germany. ³¹Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. ³²Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. ³³Clinical Research Branch, National Institute on Aging, Baltimore, Maryland 21205, USA. ³⁴Centre for Population Health Sciences, University of Edinburgh, EH8 9AG, UK. ³⁵Centre for Population Health Sciences and Institute of Genetics and Molecular Medicine, College of Medicine and Veterinary Medicine, University of Edinburgh, EH8 9AG, UK. ³⁶Croatian Centre for Global Health,

University of Split, 21000 Split, Croatia.³⁷Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, SE-751 85 Uppsala, Sweden.³⁸Amgen, 1 Kendall Square, Building 100, Cambridge, Massachusetts 02139, USA.³⁹Department of Clinical Sciences, Lund University, 205 02 Malmö, Sweden.⁴⁰Department of Medicine, University of Verona, 37134 Verona, Italy.⁴¹Ealing Hospital, London UB1 3HJ, UK.⁴²Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA.⁴³Genetic Epidemiology Group, Epidemiology and Public Health, UCL, London, WC1E 6BT, UK.⁴⁴Center for Genome Science, National Institute of Health, Seoul 122-701, Korea.⁴⁵Department of Cardiology, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands.⁴⁶Departments of Epidemiology and Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA.⁴⁷Centre for Cellular and Molecular Biology (CCMB), Council of Scientific and Industrial Research (CSIR), Uppal Road, Hyderabad 500 007, India.⁴⁸Department of Public Health and Primary Care, University of Cambridge, CB1 8RN, UK.⁴⁹Department of Basic Medical Research and Education, and Department of Geriatric Medicine, Ehime University Graduate School of Medicine, Toon, 791-0295, Japan.⁵⁰Division of Community Health Sciences, St George's University of London, London SW17 0RE, UK.⁵¹Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA.⁵²Division of Biostatistics and Department of Genetics, School of Medicine, Washington University in St. Louis, Saint Louis, Missouri 63110, USA.⁵³Department of Medicine, University of Eastern Finland and Kuopio University Hospital, 70210 Kuopio, Finland.⁵⁴Department of Preventive Medicine and Epidemiology, Loyola University Medical School, Maywood, Illinois 60153, USA.⁵⁵Department of Laboratory Medicine & Institute of Human Genetics, University of California San Francisco, 513 Parnassus Ave. San Francisco, California 94143, USA.⁵⁶Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore.⁵⁷Centre for Molecular Epidemiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore.⁵⁸Department of Clinical Chemistry, University of Tampere and Tampere University Hospital, Tampere 33521, Finland.⁵⁹Department of Genetics, Biology and Biochemistry, University of Torino, Via Santena 19, 10126 Torino, Italy.⁶⁰Human Genetics Foundation (HUGF), Via Nizza 52, 10126 Torino, Italy.⁶¹Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA.⁶²MRC Centre for Causal Analyses in Translational Epidemiology, School of Social & Community Medicine, University of Bristol, Bristol BS8 2BN, UK.⁶³Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Heidelberglaan 100, 3508 GA Utrecht, The Netherlands.⁶⁴Complex Genetics Section, Department of Medical Genetics - DBG, University Medical Center Utrecht, 3508 GA Utrecht, The Netherlands.⁶⁵Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Crawley, Western Australia 6009, Australia.⁶⁶HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, 7600 Levanger, Norway.⁶⁷MRC Unit for Lifelong Health & Ageing, London WC1B 5JU, UK.⁶⁸Occupational and Environmental Medicine, Department of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, 40530 Gothenburg, Sweden.⁶⁹Centre for Cardiovascular Genetics, University College London, London WC1E 6JF, UK.⁷⁰MRC Human Genetics Unit and Institute of Genetics and Molecular Medicine, Edinburgh EH2, UK.⁷¹Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester LE3 9QP, UK.⁷²Leicester NIHR Biomedical Research Unit in Cardiovascular Disease, Glenfield Hospital, Leicester, LE3 9QP, UK.⁷³Studies Coordinating Centre, Division of Hypertension and Cardiac Rehabilitation, Department of Cardiovascular Diseases, University of Leuven, Campus Sint Rafaël, Kapucijnenvoer 35, Block D, Box 7001, 3000 Leuven, Belgium.⁷⁴Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, Maryland 20892, USA.⁷⁵Columbia University, New York, New York 10027, USA.⁷⁶Department of Medicine III, Medical Faculty Carl Gustav Carus at the Technical University of Dresden, 01307 Dresden, Germany.⁷⁷Epidemiology and Biostatistics, School of Public Health, Imperial College, London W2 1PG, UK.⁷⁸Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, 45110 Ioannina, Greece.⁷⁹Wake Forest University Health Sciences, Winston-Salem, North Carolina 27157, USA.⁸⁰Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, USA.⁸¹Division of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, University of Texas at Houston Health Science Center, 12 Herman Pressler, Suite 453E, Houston, Texas 77030, USA.⁸²Laboratory of Neurogenetics, National Institute on Aging, Bethesda, Maryland 20892, USA.⁸³Laboratory of Cardiovascular Science, Intramural Research Program, National Institute on Aging, NIH, Baltimore, Maryland 21224, USA.⁸⁴Washington Hospital Center, Division of Cardiology, Washington, District of Columbia 20010, USA.⁸⁵ALSPAC Laboratory, University of Bristol, Bristol BS8 2BN, UK.⁸⁶Pediatric Epidemiology Center, University of South Florida, Tampa, Florida 33612, USA.⁸⁷Clinical Pharmacology Unit, University of Cambridge, Addenbrookes Hospital, Hills Road, Cambridge CB2 2QQ, UK.⁸⁸University of Dundee, Ninewells Hospital & Medical School, Dundee DD1 9SY, UK.⁸⁹Genetic Epidemiology Group, Department of Epidemiology and Public Health, UCL, London WC1E 6BT, UK.⁹⁰Pathology and Laboratory Medicine, University of Western Australia, Crawley, Western Australia 6009, Australia.⁹¹Molecular Genetics, PathWest Laboratory Medicine, Nedlands, Western Australia 6009, Australia.⁹²Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford OX3 7LF, UK.⁹³Department of Epidemiology and Biostatistics, Case Western Reserve University, 2103 Cornell Road, Cleveland, Ohio 44106, USA.⁹⁴Department of Medicine, Johns Hopkins University, Baltimore 21205, USA.⁹⁵Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, 33521, Finland.⁹⁶Department of Medicine, University of Turku and Turku University Hospital, Turku 20521, Finland.⁹⁷Department of Nutrition, University of North Carolina, Chapel Hill, North Carolina 27599, USA.⁹⁸Office of Population Studies Foundation, University of San Carlos, Talamban, Cebu City 6000, Philippines.⁹⁹Department of Neurology and Framingham Heart Study, Boston University School of Medicine, Boston, Massachusetts 02118, USA.¹⁰⁰Department of Internal Medicine II, University Medical Center Regensburg, 93053 Regensburg, Germany.¹⁰¹Department of Epidemiology and Preventive Medicine, University Medical Center Regensburg, 93053 Regensburg, Germany.¹⁰²Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland 21205, USA.¹⁰³Renal Division, University Hospital Freiburg, 79095 Freiburg, Germany.¹⁰⁴Département de Génétique Médicale, Université de Lausanne, 1015 Lausanne, Switzerland.¹⁰⁵Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.¹⁰⁶Division of Genetics, GlaxoSmithKline, Philadelphia, Pennsylvania 19101, USA.¹⁰⁷International Centre for Circulatory Health, National Heart & Lung Institute, Imperial College, London SW7 2AZ, UK.¹⁰⁸Genetics of Complex Traits, Peninsula Medical School, University of Exeter, Exeter EX4 4QJ, UK.¹⁰⁹Department of Community Health Sciences & Department of Medicine, Aga Khan University, Karachi 74800, Pakistan.¹¹⁰Medizinische Klinik II, Universität zu Lübeck, 23538 Lübeck, Germany.¹¹¹Diabetes Unit, KEM Hospital and Research Centre, Rasta Peth, Pune-411011, Maharashtra, India.¹¹²Department of Clinical Sciences, Diabetes and Endocrinology Research Unit, University Hospital, 205 02 Malmö, Sweden.¹¹³Lund University, Malmö 20502, Sweden.¹¹⁴Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02139, USA.¹¹⁵Department of Chronic Disease Prevention, National Institute for Health and Welfare, 00251 Helsinki, Finland.¹¹⁶FIMM, Institute for Molecular Medicine, Finland, Biomedicum, P.O. Box 104, 00251 Helsinki, Finland.¹¹⁷William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK.¹¹⁸Merck Research Laboratory, 126 East Lincoln Avenue, Rahway, New Jersey 07065, USA.¹¹⁹Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK.¹²⁰University of Cambridge Metabolic Research Labs, Institute of Metabolic Science Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.¹²¹Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA.¹²²Cardiovascular Genetics, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA.¹²³Department of Physiology and Biophysics, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA.¹²⁴National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.¹²⁵National Institute for Health and Welfare, 00271 Helsinki, Finland.¹²⁶Broad Institute, Cambridge, Massachusetts 02142, USA.¹²⁷Leibniz-Institute for Arteriosclerosis Research, Department of Molecular Genetics of Cardiovascular Disease, University of Münster, 48149 Münster, Germany.¹²⁸Medical Faculty of the Westfälische Wilhelms University Münster, Department of Molecular Genetics of Cardiovascular Disease, University of Münster, 48149 Münster, Germany.¹²⁹Division of Cardiology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.¹³⁰Georgia Prevention Institute, Department of Pediatrics, Medical College of Georgia, Augusta, Georgia 30912, USA.¹³¹Unit of Genetic Epidemiology and Bioinformatics, Department of Epidemiology, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands.¹³²Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA.¹³³Department of Epidemiology, University of Washington, Seattle, Washington 98195, USA.¹³⁴Group Health Research Institute, Group Health Cooperative, Seattle, Washington 98124, USA.¹³⁵Seattle Epidemiologic Research and Information Center, Veterans Health Administration Office of Research & Development, Seattle, Washington 98108, USA.¹³⁶Department of Medicine, University of Washington, Seattle, Washington 98195, USA.¹³⁷Department of Cardiology, University of Tartu, L. Puusepa 8, 51014 Tartu, Estonia.¹³⁸Tallinn University of Technology, Institute of Biomedical Engineering, Ehitajate tee 5, 19086 Tallinn, Estonia.¹³⁹Centre of Cardiology, North Estonia Medical Centre, Sõistite tee 19, 13419 Tallinn, Estonia.¹⁴⁰Department of Non-communicable disease Epidemiology, The London School of Hygiene and Tropical Medicine London, Keppel Street, London WC1E 7HT, UK.¹⁴¹South Asia Network for Chronic Disease, Public Health Foundation of India, C-1/52, SDA, New Delhi 100016, India.¹⁴²Department of Emergency and Cardiovascular Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, 41685 Gothenburg, Sweden.¹⁴³Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, 0317 Oslo, Norway.¹⁴⁴Tuscany Regional Health Agency, 50129 Florence, Italy.¹⁴⁵Tropical Medicine Research Institute, University of the West Indies, Mona, Kingston, Jamaica.¹⁴⁶University of Ibadan, 200284 Ibadan, Nigeria.¹⁴⁷Department of Genomic Medicine, and Department of Preventive Cardiology, National Cerebral and Cardiovascular Research Center, Suita, 565-8565, Japan.¹⁴⁸Department of Health Science, Shiga University of Medical Science, Otsu, 520-2192, Japan.¹⁴⁹Department of Geriatric Medicine, Osaka University Graduate School of Medicine, Suita, 565-0871, Japan.¹⁵⁰Tohoku University Graduate School of Pharmaceutical Sciences and Medicine, Sendai, 980-8578, Japan.¹⁵¹Lifestyle-related Disease Prevention Center, Shiga University of Medical Science, Otsu, 520-2192, Japan.¹⁵²Department of Medical Science and Cardiorespiratory Medicine, Yokohama City University School of Medicine, Yokohama, 236-0004, Japan.¹⁵³Department of Statistics, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago, Chile.¹⁵⁴Institute of Human Genetics, Helmholtz Zentrum Munich, German Research Centre for Environmental Health, 85764 Neuherberg, Germany.¹⁵⁵Institute of Human Genetics, Klinikum rechts der Isar, Technical University of Munich, 81675 Munich, Germany.¹⁵⁶Institute of Epidemiology, Helmholtz Zentrum Munich, German Research Centre for Environmental Health, 85764 Neuherberg, Germany.¹⁵⁷Chair of Epidemiology, Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, 81377 Munich, Germany.¹⁵⁸Klinikum Grosshadern, 81377 Munich, Germany.¹⁵⁹National Heart and Lung Institute, Imperial College London, London W12 0HS, UK.¹⁶⁰Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California 90048, USA.¹⁶¹Medical Population Genetics, Broad Institute of Harvard and MIT, 5 Cambridge Center, Cambridge, Massachusetts 02142, USA.¹⁶²National Heart, Lung and Blood Institute and its Framingham Heart Study, 73 Mount Wayte Ave., Suite #2, Framingham, Massachusetts 01702, USA.¹⁶³Department of Neurology and Medicine, University of Washington, Seattle, Washington 98195, USA.¹⁶⁴Department of Medicine (Geriatrics), University of Mississippi Medical Center, Jackson, Mississippi 39216, USA.¹⁶⁵Department of Neurology, Boston University School of Medicine, Massachusetts 02118, USA.¹⁶⁶Finnish Institute of Occupational Health, Aapistie 1, 90220 Oulu, Finland.¹⁶⁷Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.¹⁶⁸Lapland Central Hospital, Department of Physiatrics, Box 8041, 96101 Rovaniemi, Finland.¹⁶⁹Center for

Non-Communicable Diseases Karachi 74800, Pakistan.¹⁷⁰Atherosclerosis Research Unit, Department of Medicine, Karolinska Institute, 171 77 Stockholm, Sweden.¹⁷¹Department of Internal Medicine, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands.¹⁷²Department of Clinical Genetics, Erasmus Medical Center, 3000 CA Rotterdam, The Netherlands.¹⁷³Gerontology Research Center, National Institute on Aging, Baltimore, Maryland 21224, USA.¹⁷⁴Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, Cittadella Universitaria di Monserrato, 09042 Monserrato, Cagliari, Italy.¹⁷⁵Unita Operativa Semplice Cardiologia, Divisione di Medicina, Presidio Ospedaliero Santa Barbara, 09016 Iglesias, Italy.¹⁷⁶Computational Medicine Research Group, Institute of Clinical Medicine, University of Oulu and Biocenter Oulu, 90014 University of Oulu, Oulu, Finland.¹⁷⁷NMR Metabonomics Laboratory, Department of Biosciences, University of Eastern Finland, 70211 Kuopio, Finland.¹⁷⁸Department of Internal Medicine and Biocenter Oulu, Clinical Research Center, 90014 University of Oulu, Oulu, Finland.¹⁷⁹Institute for Molecular Medicine Finland FIMM, 00014 University of Helsinki, Helsinki, Finland.¹⁸⁰Department of Biomedical Engineering and Computational Science, School of Science and Technology, Aalto University, 00076 Aalto, Espoo, Finland.¹⁸¹NUS Graduate School for Integrative Sciences & Engineering (NGS) Centre for Life Sciences (CeLS), Singapore 117456, Singapore.¹⁸²Department of Internal Medicine B, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany.¹⁸³Institute of Pharmacology, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany.¹⁸⁴Institute for Community Medicine, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany.¹⁸⁵U557 Institut National de la Santé et de la Recherche Médicale, U1125 Institut National de la Recherche Agronomique, Université Paris 13, 93017 Bobigny, France.¹⁸⁶Department of Neurology, Mayo Clinic, Jacksonville, Florida 32224, USA.¹⁸⁷Imperial College Cerebrovascular Unit (ICCRU), Imperial College, London W6 8RF, UK.¹⁸⁸Faculty of Medicine, University of Split, 21000 Split, Croatia.¹⁸⁹Department of Internal Medicine, Diabetology, and Nephrology, Medical University of Silesia, 41-800, Zabrze, Poland.¹⁹⁰Public Health Sciences section, Division of Community Health Sciences, University of Edinburgh, Medical School, Teviot Place, Edinburgh, EH8 9AG, UK.¹⁹¹School of Science and Engineering, University of Ballarat, 3353 Ballarat, Australia.¹⁹²Prevention and Care of Diabetes, Department of Medicine III, Medical Faculty Carl Gustav Carus at the Technical University of Dresden, 01307 Dresden, Germany.¹⁹³University Hospital Münster, Internal Medicine D, 48149 Münster, Germany.¹⁹⁴Department of Medical Statistics, Epidemiology and Medical Informatics, Andrija Stampar School of Public Health, University of Zagreb, 10000 Zagreb, Croatia.¹⁹⁵AstraZeneca R&D, 431 83 Mölndal, Sweden.¹⁹⁶Genetic Epidemiology & Biostatistics Platform, Ontario Institute for Cancer Research, Toronto, Ontario M5G 1L7, Canada.¹⁹⁷Samuel Lunenfeld Institute for Medical Research, University of Toronto, Toronto, Ontario M5S 1A1, Canada.¹⁹⁸Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK.¹⁹⁹Department of Epidemiology and Public Health, Imperial College, Norfolk Place, London W2 1PG, UK.²⁰⁰Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and the Department of Clinical Physiology, Turku University Hospital, Turku, 20521, Finland.²⁰¹Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan Medical Center, Ann Arbor, Michigan 48109, USA.²⁰²Singapore Eye Research Institute, Singapore 168751, Singapore.²⁰³Department of Ophthalmology, National University of Singapore, Singapore 119074, Singapore.²⁰⁴Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119074, Singapore.²⁰⁵Duke-National University of Singapore Graduate Medical School, Singapore 169857, Singapore.²⁰⁶Division of Biostatistics, Washington University School of Medicine, Saint Louis, Missouri 63110, USA.²⁰⁷Department of Primary Care & Population Health, UCL, London NW3 2PF, UK.²⁰⁸BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126 University Place, Glasgow G12 8TA, UK.²⁰⁹Epidemiology Public Health, UCL, London WC1E 6BT, UK.²¹⁰Departments of Epidemiology, Biostatistics, and Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA.²¹¹Division of Nephrology, Department of Internal Medicine, University Medical Center Groningen, University of Groningen, 9713 GZ Groningen, The Netherlands.²¹²Division of Cardiology, Brigham and Women's Hospital, 900 Commonwealth Avenue East, Boston, Massachusetts 02215, USA.²¹³Geriatric Rehabilitation Unit, Azienda Sanitaria Firenze (ASF), 50100 Florence, Italy.²¹⁴National Institute for Health and Welfare, Diabetes Prevention Unit, 00271 Helsinki, Finland.²¹⁵Hjelt Institute, Department of Public Health, University of Helsinki, 00014 Helsinki, Finland.²¹⁶South Ostrobothnia Central Hospital, 60220 Seinäjoki, Finland.²¹⁷Red RECAVA Grupo RD06/0014/0015, Hospital Universitario La Paz, 28046 Madrid, Spain.²¹⁸Department of Neurology, General Central Hospital, 39100 Bolzano, Italy.²¹⁹CIBER Epidemiología y Salud Pública, 08003 Barcelona, Spain.²²⁰Department of Medicine and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.²²¹Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, Maryland 21201, USA.²²²U872 Institut National de la Santé et de la Recherche Médicale, Centre de Recherche des Cordeliers, 75006 Paris, France.²²³Institute of Physiology, Ernst-Moritz-Arndt-University Greifswald, 17487 Greifswald, Germany.²²⁴Institute of Clinical Medicine/Obstetrics and Gynecology, University of Oulu, 90014 Oulu, Finland.²²⁵Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, 1011 Lausanne, Switzerland.²²⁶Human Genetics Center, 1200 Hermann Pressler, Suite E447 Houston, Texas 77030, USA.²²⁷Division of Epidemiology and Prevention, Boston University School of Medicine, Boston, Massachusetts 02215, USA.²²⁸Department of Mathematics, Boston University, Boston, Massachusetts 02215, USA.²²⁹Institute of Health Sciences, University of Oulu, BOX 5000, 90014 University of Oulu, Finland.²³⁰Biocenter Oulu, University of Oulu, BOX 5000, 90014 University of Oulu, Finland.²³¹National Institute for Health and Welfare, Box 310, 90101 Oulu, Finland.²³²MRC-HPA Centre for Environment and Health, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK.²³³Centre of Medical Systems Biology (CMSB 1-2), NGL Erasmus Medical Center, Rotterdam, The Netherlands.

*These authors contributed equally to this work.

†A full list of authors and affiliations appears in Supplementary Information.

‡Deceased.

Dynamics of human adipose lipid turnover in health and metabolic disease

Peter Arner¹, Samuel Bernard², Mehran Salehpour³, Göran Possnert³, Jakob Liebl⁴, Peter Steier⁴, Bruce A. Buchholz⁵, Mats Eriksson¹, Erik Arner¹, Hans Hauner⁶, Thomas Skurk⁶, Mikael Rydén¹, Keith N. Frayn⁷ & Kirsty L. Spalding⁸

Adipose tissue mass is determined by the storage and removal of triglycerides in adipocytes¹. Little is known, however, about adipose lipid turnover in humans in health and pathology. To study this *in vivo*, here we determined lipid age by measuring ¹⁴C derived from above ground nuclear bomb tests in adipocyte lipids. We report that during the average ten-year lifespan of human adipocytes, triglycerides are renewed six times. Lipid age is independent of adipocyte size, is very stable across a wide range of adult ages and does not differ between genders. Adipocyte lipid turnover, however, is strongly related to conditions with disturbed lipid metabolism. In obesity, triglyceride removal rate (lipolysis followed by oxidation) is decreased and the amount of triglycerides stored each year is increased. In contrast, both lipid removal and storage rates are decreased in non-obese patients diagnosed with the most common hereditary form of dyslipidaemia, familial combined hyperlipidaemia. Lipid removal rate is positively correlated with the capacity of adipocytes to break down triglycerides, as assessed through lipolysis, and is inversely related to insulin resistance. Our data support a mechanism in which adipocyte lipid storage and removal have different roles in health and pathology. High storage but low triglyceride removal promotes fat tissue accumulation and obesity. Reduction of both triglyceride storage and removal decreases lipid shunting through adipose tissue and thus promotes dyslipidaemia. We identify adipocyte lipid turnover as a novel target for prevention and treatment of metabolic disease.

A major function of adipose tissue is to store and release fatty acids, which are incorporated into adipocyte triglycerides according to whole-body energy demands. Body fat mass is determined by the balance between triglyceride storage and removal in adipocytes, by either enzymatic hydrolysis (lipolysis) and subsequent fatty acid oxidation and/or ectopic deposition in non-adipose tissues. Little is known about the dynamics of these processes in humans. Although isotope tracer methods have been used to estimate lipid turnover in human adipose tissue, these studies have been limited to short-term experimental conditions^{1–4}. To study long-term adipose tissue lipid turnover *in vivo* and across the adult lifespan, we developed a method to retrospectively determine the age of adipocyte triglycerides in humans. Triglycerides are the major component of the adipocyte lipid droplet. Lipid age was assessed by measuring the ¹⁴C content in the lipid compartment of adipocytes from human subcutaneous adipose tissue, the major fat depot in humans. ¹⁴C levels in the atmosphere remained remarkably stable until above ground nuclear bomb tests between approximately 1955 and 1963 caused a significant increase in ¹⁴C relative to stable carbon isotope levels⁵ (Fig. 1a). After the Limited Nuclear Test Ban Treaty was signed in 1963, ¹⁴C levels in the atmosphere decreased exponentially. This is not due to radioactive decay (half-life ($T_{1/2}$) for ¹⁴C is 5,730 years), but to diffusion of ¹⁴CO₂ out of the atmosphere⁶. ¹⁴C in the atmosphere oxidises to form CO₂, which is

taken up in the biotope by photosynthesis. Because we eat plants, or animals that live off plants, the ¹⁴C content in the atmosphere is directly mirrored in the human body.

Radiocarbon dating has been used to study the incorporation of atmospheric ¹⁴C into DNA to determine the age of different human cell types, including adipocytes^{7–11}. Here, we compared the incorporation of ¹⁴C into adipocyte triglycerides with the dynamic changes in atmospheric ¹⁴C described earlier. Triglyceride age was determined by using a linear lipid replacement model in which the age distribution of lipids within an individual was exponentially distributed corresponding to a constant turnover rate (per year)¹². The associated mean age, termed lipid age, is the inverse of the turnover rate and reflects the irreversible removal of lipids from adipose stores (Supplementary Information 1 and Fig. 1 of Supplementary Information 1).

Earlier studies indicate that triglycerides in adipose tissue form two distinct pools with high or low turnover rates, respectively^{13,14}. Our data, obtained from individuals born before, during and after bomb testing, do not support the hypothesis of dual large lipid pools with different half-lives (Fig. 1b). ¹⁴C data were modelled according to one or more pools of lipids with different lipid removal rates (Supplementary Information 1). The existence of a very small pool of younger lipids cannot be excluded based on data modelling (Supplementary Information 1 and Fig. 2 of Supplementary Information 1). According to a two-pool model the influence on the turnover rate is proportional to the fraction of lipid in the small pool. Triglyceride exchange between adipocytes and other small storage pools can affect turnover estimates. The two-pools model shows, however, that the non-adipose pool can be neglected when it makes up less than 20% of the lipids (Supplementary Information 1, Fig. 3). Small pools with high turnover are more important for short-term (days or weeks) than long-term (years) triglyceride turnover.

Mean lipid age was 1.6 years (Fig. 1c), which is in the same range as in short-term turnover studies⁴. The distribution of lipid age was compared with that of adipocyte age reported previously in a comparable cohort⁹. The mean age of adipocytes was 9.5 years (Fig. 1d). This implies that triglycerides, on average, are replaced six times during the lifespan of the adipocyte, enabling a dynamic regulation of lipid storage and mobilization over time.

There is a large variation in adipocyte size within and between individuals (Supplementary Information 2, Supplementary Table 1)¹⁵. However, it is unlikely that the rate of triglyceride removal from adipocytes is important for these variations, as lipid age was not related to adipocyte size when set in relation to the body fat mass (Fig. 2a, b), nor was there a difference in lipid age between large and small adipocytes of the same adipose tissue sample (Fig. 2c, d). These data indicate that there is a continuous exchange of lipids between adipocytes within the adipose tissue that is not dependent on adipocyte size. Fatty acids produced by lipolysis in one adipocyte could, for example, be taken up

¹Department of Medicine, Karolinska University Hospital, SE-141 86 Stockholm, Sweden. ²Institut Camille Jordan, CNRS UMR 5208, University of Lyon, F-69622 Villeurbanne, France. ³Department of Physics and Astronomy, Ion Physics, Uppsala University, SE-751 20, Sweden. ⁴Faculty of Physics – Isotope Research, University of Vienna, Vienna, A-1090, Austria. ⁵Center for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratory, 7000 East Ave., L-397, Livermore, California 94551 USA. ⁶Else Kröner-Fresenius Centre for Nutritional Medicine, Technische Universität München, D-85350 Weihenstephan, Germany. ⁷Oxford Centre for Diabetes, Endocrinology & Metabolism, Churchill Hospital, Oxford OX3 7LJ, UK. ⁸Department of Cell and Molecular Biology, Karolinska Institute, SE-171 77, Stockholm, Sweden.

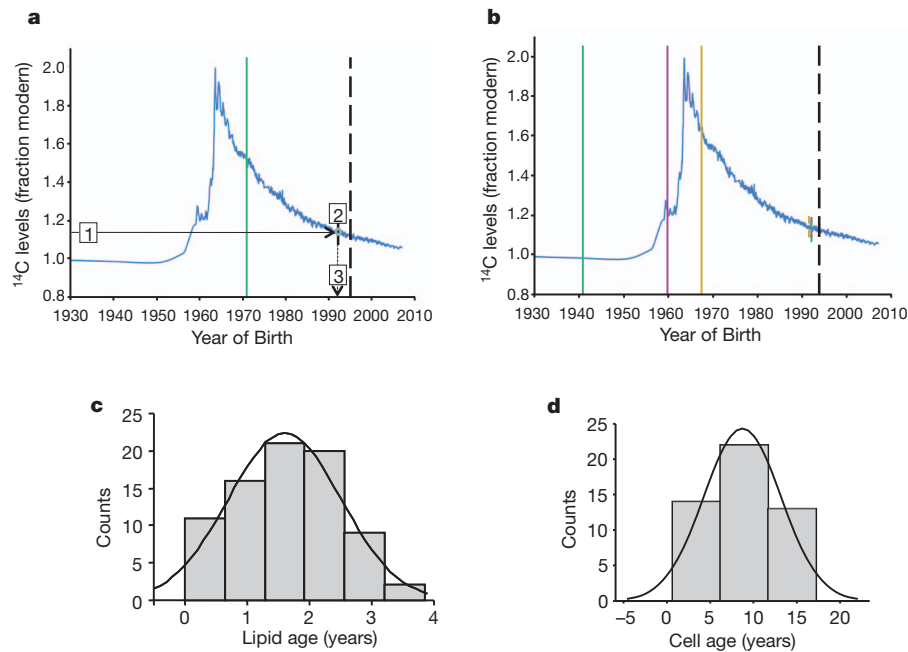


Figure 1 | Atmospheric ^{14}C over time and its use to determine lipid age and adipocyte age. **a**, Above ground nuclear bomb testing during the period of the cold war caused an increase in atmospheric levels of ^{14}C . These values decreased exponentially following implementation of a limited world-wide test ban treaty in 1963 (blue curve). Lipid age is determined by measuring ^{14}C levels in lipids (1) and plotting this value against the bomb curve (2) to determine the difference between the year corresponding to the atmospheric ^{14}C concentration (3) and the biopsy collection date (dashed line). Atmospheric ^{14}C levels are presented as $^{14}\text{C}/^{12}\text{C}$ ratios in units of fraction modern (for a definition of 'modern' see Supplementary Information 2). **b**, Lipid age and

turnover do not change as a function of person age. Lipid age is shown for three individuals born in 1940.2, 1959.9 and 1967.9. Lipid age was shown to be the same for all individuals, despite markedly different subject ages. Fat biopsies were collected from all individuals on the same date (dashed vertical line). The solid vertical lines indicate the date of birth. The small dashed lines show the ^{14}C lipid value for each individual. **c**, Distribution of values for lipid age in healthy non-obese or obese individuals from cohort 1 ($n = 78$). **d**, The distribution of values for human adipocyte age ($n = 27$). Adipocyte age data are obtained from a previous publication (see main text).

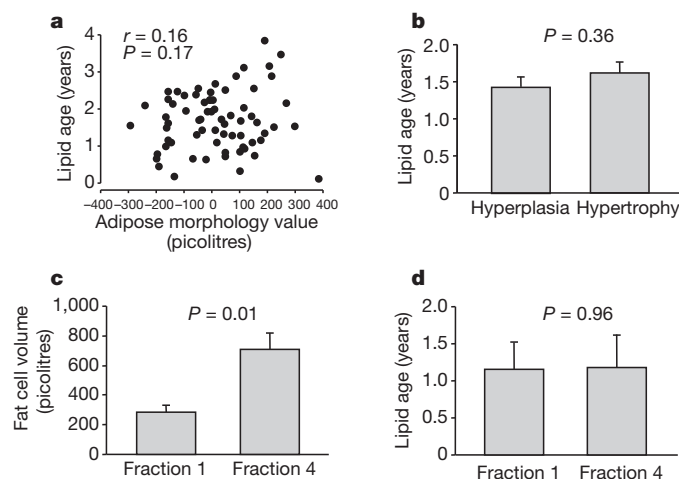


Figure 2 | Relationship between adipocyte size and lipid age. **a**, **b**, Influence of adipocyte cellularity on lipid age. Individuals were assigned a morphometric value, which is the difference between the measured adipocyte volume for the individual minus the average adipocyte volume for all subjects (see Supplementary Information 2). This analysis was carried out across the full range of body masses. Positive values indicate larger adipocytes than expected (fewer but larger adipocytes = hypertrophy). Negative values indicate smaller adipocytes than expected (many but smaller adipocytes = hyperplasia). **a**, Individual values compared by linear regression analysis ($n = 74$). **b**, Data (mean \pm standard error) with morphology as a dichotomous variable ($n = 36$ for hyperplasia and $n = 38$ for hypertrophy). An unpaired t -test was used. **c**, **d**, Isolated subcutaneous adipocytes were fractioned into small (fraction 1) or very large (fraction 4) samples ($n = 7$). Adipocyte volume (**c**) and adipose lipid age (**d**) were compared. Values are mean \pm standard error. A paired t -test was used. Data in **a** and **b** are from non-obese plus obese individuals in cohort 1 and data in **c** and **d** are from cohort 2.

by adjacent adipocytes and incorporated into their triglycerides. These processes would not be part of lipid removal as measured here.

Lipid age and total fat mass data were used to determine the net triglyceride storage in adipose tissue (kg year^{-1}) (see Supplementary Information 1). The net amount of lipid stored in adipose tissue each year is the sum of exogenous fat incorporation and endogenous synthesis, minus lipid removal. The removal rate represents the hydrolysis of triglycerides (lipolysis) followed by the irreversible removal of lipids by oxidation. A high lipid age therefore mirrors low removal rates. No relationship between lipid storage or removal and person age or gender was seen (Supplementary Information 2 and Fig. 1a–d of Supplementary Information 2).

Two clinical conditions where altered lipid metabolism is observed were investigated—obesity and familial combined hyperlipidaemia (FCHL); the latter is the most common hereditary lipid disorder (reviewed in ref. 16). It has an unknown aetiology and is a common hereditary cause of premature coronary heart disease. Adipocyte lipolysis is impaired in both conditions due to decreased cyclic AMP-dependent signalling, the major lipolytic pathway in adipocytes^{17–19}. Both conditions show a similar metabolic phenotype (mixed dyslipidaemia, elevated apolipoprotein B and insulin resistance)²⁰. These clinical characteristics are confirmed in our study cohort (Supplementary Information 2, Supplementary Table 1). FCHL individuals may present with a range of body fat levels; however, for our analyses only non-obese FCHL patients were selected so as to remove the confounding factor of obesity from the study.

In obese subjects, the rate of triglyceride storage (Fig. 3a) and mean lipid age (Fig. 3b) were markedly increased compared to non-obese individuals. Both lipid age ($r = 0.38$, $P = 0.0005$) and triglyceride storage ($r = 0.60$, $P < 0.0001$) correlated with body mass index (BMI) when non-obese and obese individual were pooled together. Similarly, in non-obese FCHL individuals lipid age was increased to values

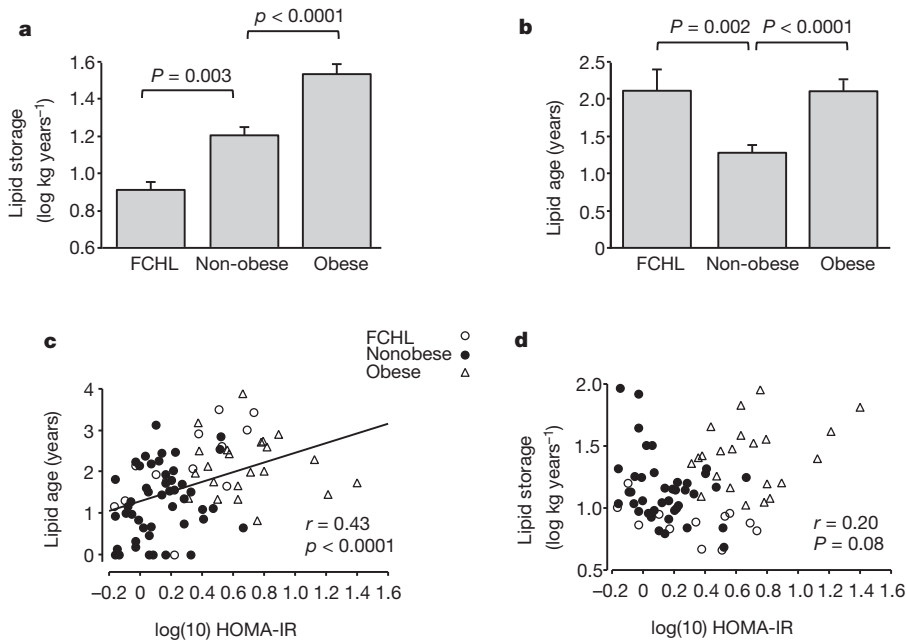


Figure 3 | Lipid turnover in subcutaneous fat. **a, b**, Lipid storage (**a**) and lipid age (**b**) were determined in 48 non-obese, 30 obese and 13 non-obese FCHL subjects. Error bars indicate standard error. Overall effect is $P < 0.0001$ by analysis of variance (ANOVA) in **a** and **b**. Results in graphs are from post-hoc test. Data are from cohort 1 (see Supplementary Information 2). A linear regression analysis was performed on all individuals from cohort 1 having

insulin resistance measures ($n = 82$). **c, d**, HOMA-IR was correlated with lipid age (**c**) and lipid storage (**d**). The relationship between lipid age and HOMA-IR remained significant when BMI, gender or group (non-obese, obese, FCHL) were included in the analysis (partial $r = 0.41$, $P = 0.006$ with BMI using multiple regression analysis and $F = 16.6$, $P = 0.0001$ and $F = 4.8$, $P = 0.03$ for gender or group, respectively, using analysis of covariance (ANCOVA)).

observed in obesity (Fig. 3b). In contrast to obesity, however, the rate of triglyceride storage was markedly decreased compared to non-obese individuals (Fig. 3a). Thus, adipocyte triglyceride turnover is not just a mere reflection of the fat mass. Our data indicate a model where a combination of high storage and low lipid removal rates, as in obesity, facilitates triglyceride accumulation within adipose tissue, thereby promoting the development and/or maintenance of excess body fat mass.

Conversely, a low rate of both triglyceride storage and removal, as in FCHL, leads to reduced triglyceride turnover and thereby a decreased ability of adipocytes to store and release fatty acids, despite a normal body fat mass. As discussed in detail elsewhere^{21,22}, low lipid turnover in adipose tissue may result in fatty acids being shunted to the liver, which drives the synthesis of apolipoprotein B and increases the circulating levels of triglycerides. Adipocyte triglyceride turnover may also be

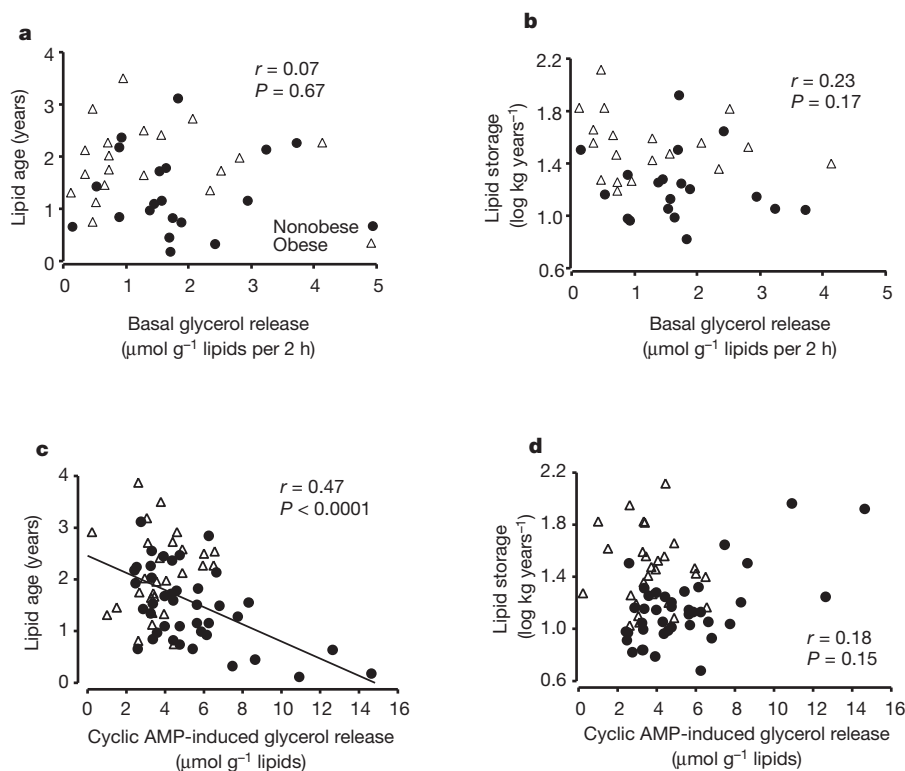


Figure 4 | Correlation between lipid turnover and adipocyte lipolysis. **a–d**, Lipid age and lipid removal rates were compared with basal rate of glycerol release (**a, b**) and with the rate of glycerol release induced with dibutyryl cyclic AMP (**c, d**), which is a phosphodiesterase-resistant and stable cyclic AMP analogue stimulating the protein kinase A complex. Linear regression analysis was used. Data are from non-obese ($n = 48$) and obese ($n = 28$) individuals from cohort 1. Data with dibutyryl cyclic AMP-induced lipolysis versus lipid age were also significant when analysed using BMI as a covariate in multiple regression analysis (partial $r = -0.40$; $P = 0.0006$).

involved in determining overall insulin effects. Insulin resistance (indirectly measured by the HOMA-IR index, see Supplementary Information 2) and lipid turnover were assessed in 82 individuals. Triglyceride age was strongly related to levels of insulin resistance (Fig. 3c), although there was no relationship between triglyceride storage and insulin resistance (Fig. 3d). There was no significant interaction between groups (lean, obese and non-obese FCHL) as determined by analysis of co-variance, indicating that the rate of triglyceride removal from adipocytes has an impact on whole-body insulin sensitivity independent of any underlying disorder. Multiple regression analysis showed that the relationship between HOMA-IR and lipid removal was not influenced by plasma triglycerides (partial $r = 0.35$; $P = 0.007$).

We also examined non-obese and obese individuals separately (Supplementary Information 2 and Figure 2a–d of Supplementary Information 2). Variations in BMI were significantly related to lipid age only among non-obese and to lipid storage only among obese individuals. HOMA-IR variations were significantly related to lipid storage when obese subjects were removed from the analysis (no relationship was found among obese subjects themselves). Thus, variations in triglyceride turnover may have a different impact on metabolic status in obese versus non-obese populations. Clearly, this assumption must be confirmed by investigations in much larger samples.

Because adipose tissue lipolysis is the first step in lipid removal, we investigated the ability of the cyclic AMP system to activate lipolysis *in vitro* in adipocytes isolated from lean and obese individuals and compared this with *in vivo* measurements of lipid storage and removal. Spontaneous (basal) lipolysis was not related to lipid turnover (Fig. 4a, b). However, the stimulated rate of lipolysis was positively correlated with triglyceride removal (inversely correlated with lipid age) but was not related to the rate of triglyceride uptake (lipid storage). This was irrespective of whether lipolysis was induced using a cyclic AMP analogue (Fig. 4c, d), by activating endogenous adenylate cyclase (using forskolin; Supplementary Information 2 and Fig. 3a, b of Supplementary Information 2) or by administration of a synthetic β -adrenoceptor-selective catecholamine (isoprenaline; Fig. 3c, d of Supplementary Information 2). These data indicate that lipolysis determines lipid turnover in adipocytes by regulating the rate of triglyceride removal. The impact of subsequent fatty acid oxidation could not be examined in this study; however, decreased lipid oxidation is frequently observed in obesity^{23,24}. As there are regional variations in lipolysis and all our studies were performed on one fat depot no attempts were made to extrapolate findings to the whole-body level.

We are in the midst of a global epidemic of obesity with negative health and socio-economic consequences. We propose adipose triglyceride turnover as a novel target for the prevention and treatment of excess body fat and possibly its consequences for insulin resistance. New insights into abnormal triglyceride turnover in FCHL patients may also suggest novel treatment strategies for this complex disease that targets adipocytes.

METHODS SUMMARY

Subjects. Subcutaneous adipose tissue was obtained from two patient cohorts. Patient selection and collection of clinical data are described in Supplementary Information 2.

Preparation of lipids. Triglycerides were extracted from pieces of adipose tissue or isolated adipocytes. Details of lipid extraction and adipocyte isolation are given in Supplementary Information 2. Extracted lipids were subjected to accelerator mass spectrometry analysis, as described in Supplementary Information 2.

Data analysis. Calculations between lipid turnover and clinical or adipocyte phenotypes are described in Supplementary Information 2. Calculations of lipid age and net lipid uptake by adipose tissue are described in Supplementary Information 1. Conventional statistical methods were used to summarize and compare data.

Received 24 June; accepted 5 August 2011.

Published online 25 September 2011.

1. Klein, R. A., Halliday, D. & Pittet, P. G. The use of 13-methyltetradecanoic acid as an indicator of adipose tissue turnover. *Lipids* **15**, 572–579 (1980).

2. Mårin, P., Oden, B. & Björntorp, P. Assimilation and mobilization of triglycerides in subcutaneous abdominal and femoral adipose tissue *in vivo* in men: effects of androgens. *J. Clin. Endocrinol. Metab.* **80**, 239–243 (1995).
3. Mårin, P., Rebuffe-Scrive, M. & Björntorp, P. Uptake of triglyceride fatty acids in adipose tissue *in vivo* in man. *Eur. J. Clin. Invest.* **20**, 158–165 (1990).
4. Strawford, A., Antelo, F., Christiansen, M. & Hellerstein, M. K. Adipose tissue triglyceride turnover, *de novo* lipogenesis, and cell proliferation in humans measured with $2\text{H}_2\text{O}$. *Am. J. Physiol.* **286**, E577–E588 (2004).
5. Nydal, R. & Lovseth, K. Distribution of radiocarbon from nuclear tests. *Nature* **206**, 1029–1031 (1965).
6. Levin, I. & Kromer, B. The tropospheric $^{14}\text{CO}_2$ level in mid latitudes of the northern hemisphere (1959–2003). *Radiocarbon* **46**, 1261–1272 (2004).
7. Spalding, K. L., Bhardwaj, R. D., Buchholz, B. A., Druid, H. & Frisen, J. Retrospective birth dating of cells in humans. *Cell* **122**, 133–143 (2005).
8. Perl, S. *et al.* Significant human β -cell turnover is limited to the first three decades of life as determined by *in vivo* thymidine analog incorporation and radiocarbon dating. *J. Clin. Endocrinol. Metab.* **95**, E234–E239 (2010).
9. Spalding, K. L. *et al.* Dynamics of fat cell turnover in humans. *Nature* **453**, 783–787 (2008).
10. Bergmann, O. *et al.* Evidence for cardiomyocyte renewal in humans. *Science* **324**, 98–102 (2009).
11. Bhardwaj, R. D. *et al.* Neocortical neurogenesis in humans is restricted to development. *Proc. Natl Acad. Sci. USA* **103**, 12564–12568 (2006).
12. Bernard, S., Frisen, J. & Spalding, K. L. A mathematical model for the interpretation of nuclear bomb test derived ^{14}C incorporation in biological systems. *Nucl. Instrum. Meth. B* **268**, 1295–1298 (2010).
13. Ekstedt, B. & Olivecrona, T. Uptake and release of fatty acids by rat adipose tissue: last in to first out? *Lipids* **5**, 858–860 (1970).
14. Kerpel, S., Shafir, E. & Shapiro, B. Mechanism of fatty acid assimilation in adipose tissue. *Biochim. Biophys. Acta* **46**, 495–504 (1961).
15. Björntorp, P. Effects of age, sex, and clinical conditions on adipose tissue cellularity in man. *Metabolism* **23**, 1091–1102 (1974).
16. Grundy, S. M., Chait, A. & Brunzell, J. D. Familial combined hyperlipidemia workshop. *Arterioscler. Thromb. Vasc. Biol.* **7**, 203–207 (1987).
17. Langin, D. *et al.* Adipocyte lipases and defect of lipolysis in human obesity. *Diabetes* **54**, 3190–3197 (2005).
18. Reynisdottir, S., Eriksson, M., Angelin, B. & Arner, P. Impaired activation of adipocyte lipolysis in familial combined hyperlipidemia. *J. Clin. Invest.* **95**, 2161–2169 (1995).
19. van der Kallen, C. J. *et al.* Evidence of insulin resistant lipid metabolism in adipose tissue in familial combined hyperlipidemia, but not type 2 diabetes mellitus. *Atherosclerosis* **164**, 337–346 (2002).
20. Ayyobi, A. F. & Brunzell, J. D. Lipoprotein distribution in the metabolic syndrome, type 2 diabetes mellitus, and familial combined hyperlipidemia. *Am. J. Cardiol.* **92**, 27–33 (2003).
21. Arner, P. Is familial combined hyperlipidaemia a genetic disorder of adipose tissue? *Curr. Opin. Lipidol.* **8**, 89–94 (1997).
22. de Graaf, J., Veerkamp, M. J. & Stalenhoef, A. F. Metabolic pathogenesis of familial combined hyperlipidaemia with emphasis on insulin resistance, adipose tissue metabolism and free fatty acids. *J. R. Soc. Med.* **95** (suppl. 42), 46–53 (2002).
23. Blaak, E. E. *et al.* Fat oxidation before and after a high fat load in the obese insulin-resistant state. *J. Clin. Endocrinol. Metab.* **91**, 1462–1469 (2006).
24. Houmard, J. A. Intramuscular lipid oxidation and obesity. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **294**, R1111–R1116 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This study was supported by the Swedish Research Council, Swedish Foundation for Strategic Research, Swedish Heart and Lung Foundation, Novo Nordic Foundation, Swedish Diabetes Foundation, Strategic Research Program in Diabetes at the Karolinska Institutet, Swedish Cancer Society, Uppsala BIO, Sweden, NIH/NCRR Grant RR13461, ERC grant 261258-HUFATREG and by the projects 'Hepatic and adipose tissue and functions in the metabolic syndrome' (HEPADIP, <http://www.hepadip.org/>) and 'Adipokines as drug targets to combat adverse effects of excess adipose tissue' (ADAPT, <http://www.adapt-eu.net/>), which were supported by the European Commission as an Integrated Project under the 6th and the 7th Framework Programmes (contract LSHM-CT-2005-018734 and contract HEALTH-F2-2008-201100). This work was performed in part under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. The authors would like to acknowledge E. Sjölén, K. Wåhlén, B.-M. Leijonhufvud, K. Hertel and Y. Widlund for technical assistance. We would like to thank F. Barnabé-Heider and J. Frisén for useful comments on the manuscript.

Author Contributions K.L.S. and P.A. designed the study and wrote the manuscript together with K.N.F. and S.B. M.R. co-ordinated writing and data assembly. S.B. and E.A. were responsible for the modelling. K.L.S. performed sample preparation. M.S., G.P., B.A.B., P.S. and J.L. performed ^{14}C accelerator mass spectrometry measurements. P.A., M.E., T.S. and H.H. collected clinical material.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.A. (peter.arnar@ki.se) or K.L.S. (kirsty.spalding@ki.se).

Endonuclease G is a novel determinant of cardiac hypertrophy and mitochondrial function

Chris McDermott-Roe¹, Junmei Ye², Rizwan Ahmed¹, Xi-Ming Sun¹, Anna Serafin³, James Ware¹, Leonardo Bottolo¹, Phil Muckett¹, Xavier Cañas³, Jisheng Zhang², Glenn C. Rowe⁴, Rachel Buchan¹, Han Lu¹, Adam Braithwaite¹, Massimiliano Mancini⁵, David Hauton⁶, Ramon Martí⁷, Elena García-Arumi⁷, Norbert Hubner^{8,9}, Howard Jacob¹⁰, Tadao Serikawa¹¹, Vaclav Zidek¹², Frantisek Papousek¹², Frantisek Kolar¹², Maria Cardona², Marisol Ruiz-Meana¹³, David García-Dorado¹³, Joan X. Comella^{14,15}, Leanne E. Felkin¹⁶, Paul J. R. Barton^{16,17}, Zoltan Arany⁴, Michal Pravenec¹², Enrico Petretto^{1,18}, Daniel Sanchis² & Stuart A. Cook^{1,17}

Left ventricular mass (LVM) is a highly heritable trait¹ and an independent risk factor for all-cause mortality². So far, genome-wide association studies have not identified the genetic factors that underlie LVM variation³, and the regulatory mechanisms for blood-pressure-independent cardiac hypertrophy remain poorly understood^{4,5}. Unbiased systems genetics approaches in the rat^{6,7} now provide a powerful complementary tool to genome-wide association studies, and we applied integrative genomics to dissect a highly replicated, blood-pressure-independent LVM locus on rat chromosome 3p. Here we identified endonuclease G (*Endog*), which previously was implicated in apoptosis⁸ but not hypertrophy, as the gene at the locus, and we found a loss-of-function mutation in *Endog* that is associated with increased LVM and impaired cardiac function. Inhibition of *Endog* in cultured cardiomyocytes resulted in an increase in cell size and hypertrophic biomarkers in the absence of pro-hypertrophic stimulation. Genome-wide network analysis unexpectedly implicated *ENDOG* in fundamental mitochondrial processes that are unrelated to apoptosis. We showed direct regulation of *ENDOG* by *ERR-α* and *PGC1α* (which are master regulators of mitochondrial and cardiac function)^{9–11}, interaction of *ENDOG* with the mitochondrial genome and *ENDOG*-mediated regulation of mitochondrial mass. At baseline, the *Endog*-deleted mouse heart had depleted mitochondria, mitochondrial dysfunction and elevated levels of reactive oxygen species, which were associated with enlarged and steatotic cardiomyocytes. Our study has further established the link between mitochondrial dysfunction, reactive oxygen species and heart disease and has uncovered a role for *Endog* in maladaptive cardiac hypertrophy.

Increased LVM is a clinically important trait that independently predicts the risk of heart failure, sudden death and all-cause mortality². Although LVM is a heritable complex trait¹, large genome-wide association studies have not identified LVM-associated genes³. Blood-pressure-dependent regulation of LVM, which is perhaps surprisingly limited⁷, has been studied extensively in model systems and acts through well-characterized and overlapping signalling modules¹². By contrast, the pathways that underlie blood-pressure-independent cardiac

hypertrophy, which is commonly seen in obesity and type 2 diabetes and is associated with mitochondrial dysfunction and lipotoxicity^{4,5}, remain largely unknown. Here we took advantage of the recent step changes in integrative systems genetics approaches in the rat^{6,7} to dissect a blood-pressure-independent cardiac mass quantitative trait locus (QTL) and to identify the causative gene and underlying mechanism.

The rat is unique for the study of cardiac mass, with more than 75 QTLs identified for this trait (Rat Genome Database; <http://rgd.mcw.edu/>). Rat chromosome 3p (0–25 megabase pairs (Mbp)) contains a highly replicated and blood-pressure-independent QTL for cardiac mass, which has been mapped in crosses of the spontaneously hypertensive rat (SHR) or the SHR stroke prone (SHRSP) rat to Wistar Kyoto (WKY) or salt sensitive (SS) rats^{13,14}. To dissect this locus genetically, we generated an F₂ intercross of SHR and Brown Norway (BN) strains and further replicated the LVM QTL (logarithm of odds (LOD) = 4.2) (Fig. 1a). We confirmed the blood-pressure-independent QTL effect in a congenic strain (SHR.BN-(3L)) that had a lower LVM and smaller cardiomyocytes than the SHR strain (Fig. 1b, c), and we refined the QTL region (to chromosome 3, 6.4–11.2 Mbp) using a second congenic strain (SHR.BN-(3S)) (Supplementary Fig. 1). In the F₂ cross, in the SHR.BN-(3L) strain and in previous experimental crosses^{13,14}, the SHR allele at the locus was associated with increased cardiac mass, and this effect was blood pressure independent (Fig. 1a, b, d). Functional assessment *in vivo* revealed that the SHR.BN-(3L) strain had better cardiac performance at baseline and after stimulation, compared with the SHR strain (Supplementary Fig. 1). These data show that an SHR allele at the cardiac mass QTL on rat chromosome 3p increases LVM and adversely affects cardiac function.

We used the new genotypes generated in our F₂ cross and those from previous experiments^{13,14} to refine the QTL region, and we identified five distinct loci (spanning 750 kilobase pairs in total) that co-segregated with the haplotypes associated with LVM variation (Fig. 1e). *Endog*, which we had previously shown to be *cis* regulated in the heart ($P = 3 \times 10^{-6}$)⁷, was the only gene at these loci that was differentially regulated in a consistent direction in the SHR and SHRSP hearts compared with the WKY heart (Supplementary Table 1). *ENDOG* is a nuclear-encoded, mitochondria-localized nuclease with

¹Medical Research Council Clinical Sciences Centre, Faculty of Medicine, Imperial College London, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK. ²Cell Signaling & Apoptosis Group, University Lleida, Biomedical Research Institute of Lleida (IRBLLEIDA), Avenida Rovira Roure 80, 25198 Lleida, Spain. ³Platform of Applied Research on Laboratory Animal, Barcelona Science Park, Baldori Reixac 4, 08028 Barcelona, Spain. ⁴Cardiovascular Institute, Beth Israel Deaconess Medical Institute, CLS906, 3 Blackfan Circle, Boston, Massachusetts 02215, USA. ⁵Department of Radiological, Oncological and Anatomic-Pathological Sciences, University of Rome, 00161 Sapienza, Italy. ⁶School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. ⁷Unitat de Patologia Mitochondrial i Neuromuscular CIBERER, Institut de Recerca Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, 08035 Barcelona, Spain. ⁸Max-Delbrück Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany. ⁹CC4, Campus Charité Mitte, Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. ¹⁰Department of Physiology, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, USA. ¹¹Institute of Laboratory Animals, Graduate School of Medicine, Kyoto University, Yoshidakonoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. ¹²Institute of Physiology, Academy of Sciences of the Czech Republic, Videnska 1083, 142 20 Prague 4, Czech Republic. ¹³Grup de Patologia Cardiovascular, Institut de Recerca Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Pg. Vall d'Hebron, 119, 08035 Barcelona, Spain. ¹⁴Cell Signaling & Apoptosis Group at CIBERNED and Vall d'Hebron Institute of Research (VHIR), Pg. Vall d'Hebron, 119-129, 08035 Barcelona, Spain. ¹⁵Department of Biochemistry and Molecular Biology and the Institut de Neurociències at Universitat Autònoma de Barcelona, 08035 Barcelona, Spain. ¹⁶Heart Science Centre, National Heart and Lung Institute, Imperial College London, Harefield Hospital, Harefield, Middlesex UB9 6JH, UK. ¹⁷Cardiovascular Biomedical Research Unit, Royal Brompton and Harefield NHS Trust, Sydney Street, London SW3 6NP, UK. ¹⁸Department of Epidemiology and Biostatistics, Faculty of Medicine, Imperial College London, Praed Street, London W2 1PG, UK.

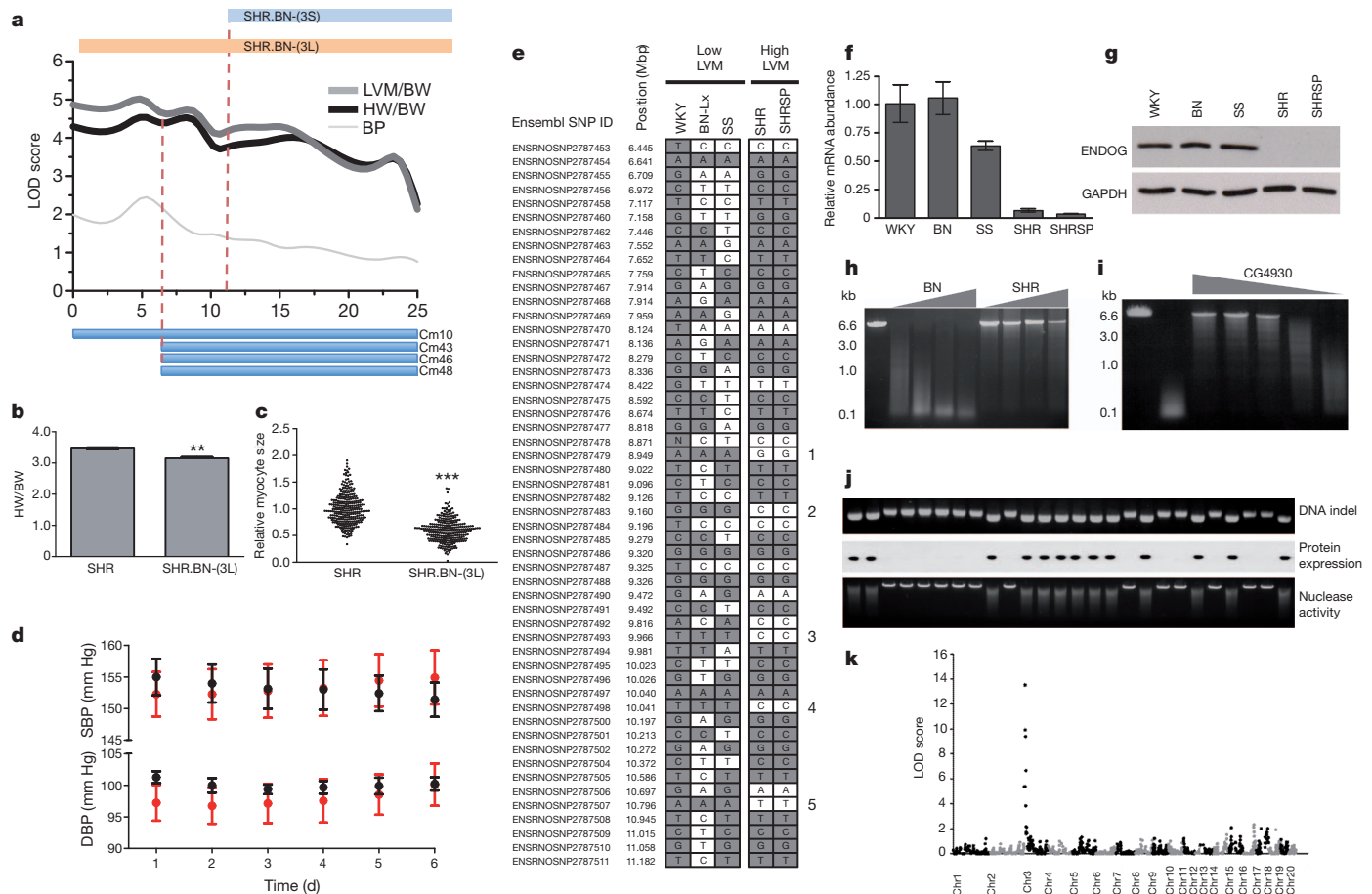


Figure 1 | Positional cloning of *Endog* as the gene underlying the rat chromosome 3p cardiac mass QTL. **a**, Mapping of heart weight (HW) and LVM corrected for body weight (BW) to chromosome 3p in the BN × SHR F₂ population. The telomeric limits of the congenic strains (SHR.BN-(3L) and SHR.BN-(3S)) and the previously mapped cardiac mass (Cm) QTLs^{13,14} are shown. The x axis indicates the physical position in Mb, and the dashed lines show the limits of the refined QTL. BP, blood pressure. **b**, HW indexed to BW in the SHR (*n* = 4) and the SHR.BN-(3L) congenic strains (*n* = 5). **c**, Relative cardiomyocyte cross-sectional area in SHR and SHR.BN-(3L) congenic strains. **d**, *In vivo* telemetric systolic blood pressure (SBP) and diastolic blood pressure (DBP) measurements in the SHR (red) and SHR.BN-(3L) (black) strains (*n* = 8 per genotype). d, days. **e**, Haplotype analysis of the refined QTL region. Single nucleotide polymorphisms (SNPs) are depicted with reference to WKY alleles (grey, identical; and white, dissimilar) with numbers (1–5) denoting the regions that are polymorphic between strains with high LVM and low LVM. BN-Lx, Brown Norway-Lx; N, unidentified nucleotide. **f**, **g**, Quantitative PCR of *Endog*

messenger RNA expression (**f**) and immunoblotting analysis of ENDOG protein expression (**g**) in strains with low or high cardiac mass at the chromosome 3p locus. GAPDH is a loading control. **h**, Nuclease activity in BN and SHR heart extracts over a range of cardiac protein extract amounts (grey wedge) (see Supplementary Methods). The first lane shows linearized plasmid. **i**, Reversal of nuclease activity in BN-Lx cardiac lysates by addition of a *Drosophila*-derived inhibitor of ENDOG¹⁸, CG4930 (range 1,500–1.5 nM, grey wedge). The first lane shows linearized plasmid, and the second lane shows cardiac lysate in the absence of inhibitor. **j**, Association of the *Endog* indel (insertion and/or deletion) with loss of ENDOG protein expression and diminished nuclease activity in the recombinant inbred strains. Top, centre and bottom panels show the DNA indel, protein expression and nuclease activity, respectively. **k**, Linkage mapping of nuclease activity in the recombinant inbred strains using a quantitative fluorescence-based assay (see Supplementary Methods). Chr, chromosome. All data are represented as mean ± s.e.m. *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001.

a proposed but disputed function in apoptosis^{8,15–17} and no known effect on cardiac mass or function. We observed reduced expression of *Endog* transcripts and lack of ENDOG protein in all strains that had increased cardiac mass (Fig. 1f, g). Sequencing of *Endog* revealed promoter and coding sequence variation, and we identified an SHR-specific, frame-shift-causing insertion in exon 1 of *Endog* that was associated with increased heart weight and LVM (Supplementary Fig. 2). There was a marked reduction in cardiac nuclease activity, which was ENDOG dependent¹⁸, in the SHR heart compared with the BN heart (Fig. 1h, i). In recombinant inbred strains derived from the SHR and BN strains^{6,7}, we confirmed the direct relationship between the insertion in SHR and the lack of nuclease activity (Fig. 1j), and we mapped *Endog*-dependent nuclease activity to a single locus that encodes *Endog* (Fig. 1k). These data identify *Endog* as the candidate gene at the QTL and implicate *Endog* loss of function as the mechanism for increased cardiac mass and impaired heart function.

We performed immunoblotting across rat and mouse tissues and determined that ENDOG was most highly expressed in the heart, where it was localized to cardiomyocytes (Fig. 2a, b) and co-localized with mitochondria (Supplementary Fig. 3). Using a short hairpin RNA (shRNA) knockdown of *Endog* (sh*Endog*)¹⁹, we tested the effect of *Endog* loss of function in cardiomyocytes and observed an increase in hypertrophic biomarkers and cell size in the absence of pro-hypertrophic stimulation (Fig. 2c, d). Conventional blood-pressure-dependent hypertrophic signalling pathways¹² were not activated in sh*Endog*-treated cells, but AMP-activated protein kinase (AMPK) was activated (Supplementary Fig. 4), which can induce cardiac hypertrophy²⁰. We also observed increased amounts of reactive oxygen species (ROS), which are also pro-hypertrophic stimuli²¹ that act through multiple downstream effectors (Supplementary Fig. 4). These data show that *Endog* loss of function directly induces cardiomyocyte hypertrophy *in vitro* and that this hypertrophy is associated with the activation of two

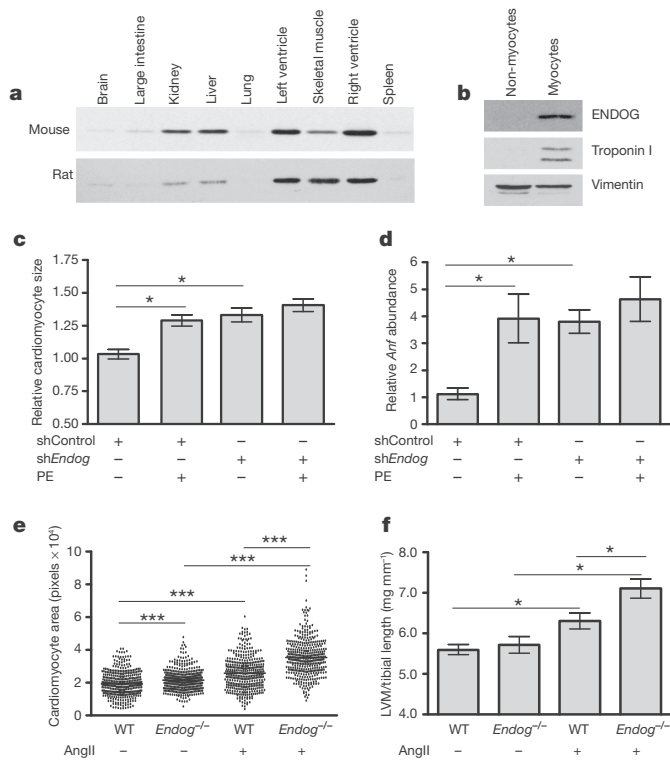


Figure 2 | *Endog* regulates cardiac hypertrophy. **a**, Immunoblotting analysis of ENDOG expression in mouse and rat tissues (ENDOG, ~30 kDa).

b, Immunoblotting analysis of ENDOG expression in cardiomyocyte and non-cardiomyocyte populations isolated from neonatal rat hearts. **c**, The size of cardiomyocytes ($n \geq 100$ cells, $n = 3$ independent experiments) treated with shRNA against *Endog* (sh*Endog*) or control shRNA (shControl) in the presence or absence of the hypertrophic stimulant phenylephrine (PE, 100 μ M, 24 h). **d**, Expression of the hypertrophic biomarker *Anrf* (which encodes atrial natriuretic factor) in sh*Endog*- or shControl-treated cardiomyocytes.

e, Cardiomyocyte size (see also Supplementary Fig. 5) in *Endog*^{-/-} and wild-type (WT) mice at baseline and after angiotensin II (AngII)-induced cardiac hypertrophy. **f**, LVM to tibial length ratio in *Endog*^{-/-} and WT mice at baseline and after stimulation with AngII. **c–f**, Data are presented as mean \pm s.e.m. *, $P < 0.05$; ***, $P < 0.001$.

pro-hypertrophic pathways, both of which have previously been linked to mitochondrial dysfunction^{20–22}.

We then examined the effects of *Endog* loss of function *in vivo* in the *Endog*-deleted (*Endog*^{-/-}) mouse¹⁷, which shows no detectable difference in apoptotic phenotypes compared with wild-type mice, an observation that was confirmed in an independent *Endog*-deleted strain¹⁶. Compared with controls, *Endog*^{-/-} mice had larger cardiomyocytes at baseline (Fig. 2e) in the absence of stimulation, in keeping with our observations in the SHR.BN-(3L) rat (Fig. 1c) and *in vitro* (Fig. 2c). Following angiotensin-II-mediated stimulation of hypertrophy, which is largely ROS dependent²¹, we observed an increase in cardiomyocyte size, hypertrophic biomarker expression and LVM in *Endog*^{-/-} mice (Fig. 2e, f and Supplementary Fig. 5). *Endog*^{-/-} mice had blood pressures that were equivalent to those of control mice at baseline ($P = 0.49$) and after stimulation with angiotensin II ($P = 0.51$) (data not shown). Together, our *in vitro* and *in vivo* data confirm a role for *Endog* in cardiomyocyte hypertrophy and identify ROS as conserved pro-hypertrophic stimuli in both systems.

Endog has been proposed to be important for apoptotic cell death⁸; however, this was subsequently disputed^{16,17}, and it was unclear how *Endog* loss of function was associated with cardiac hypertrophy and dysfunction. To infer the function of *ENDOG* in the human heart, we carried out genome-wide co-expression network analysis²³ in a large human cardiac expression data set ($n = 210$) (see Supplementary Methods). *ENDOG* was identified in a network that was highly

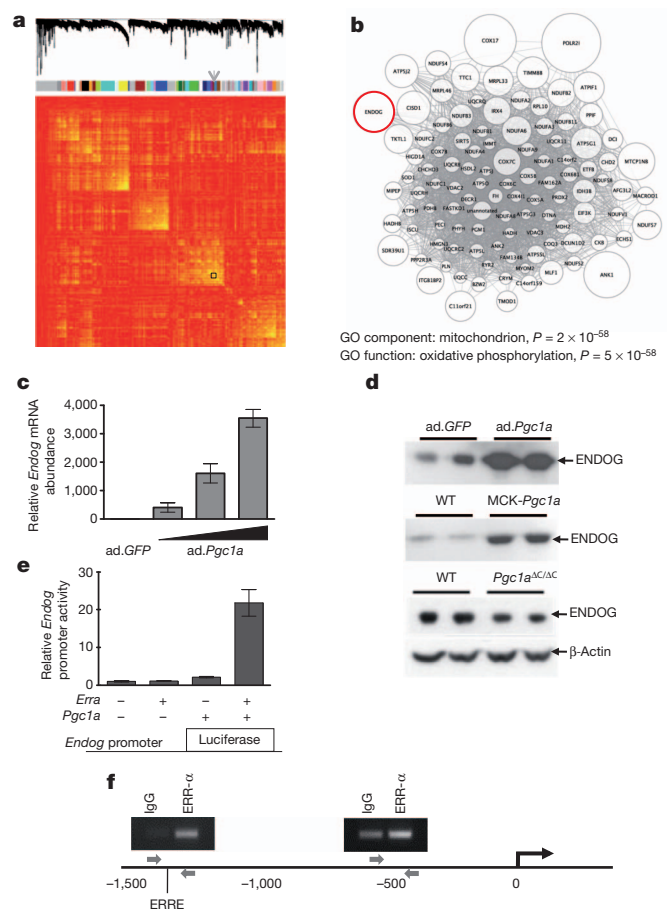


Figure 3 | *ENDOG* is co-expressed with a mitochondria-specific gene network and is regulated by PGC1 α and ERR- α . **a**, The genes (8,490 from 210 data sets) were clustered and plotted based on the dissimilarity metric between their expression profiles (see Supplementary Methods). Low-hanging branches in the dendrogram (top) represent groups of genes (modules) that have a high similarity metric. Modules are shown beneath the dendrogram (centre) and are colour coded. The arrow indicates the module (also boxed) that contains *ENDOG*. In the heat map of the correlations between expression profiles (bottom), high and low similarities are coloured yellow and red, respectively. **b**, Weighted gene co-expression network analysis (WGCNA)²³ for the module that contains *ENDOG*, providing functional annotation through cellular localization by Gene Ontology (GO) classification (Supplementary Tables 2 and 3). Nodes represent genes, and edges represent significant co-expression between genes. The node size is proportional to the relative degree of interconnectivity of each gene within the module. **c**, Quantitative PCR analysis of *Endog* expression in cultured cardiomyocytes after infection with adenovirus (ad) expressing green fluorescent protein (GFP) (ad.GFP) or increasing amounts of adenovirus expressing *Pgc1a* (ad.*Pgc1a*) (wedge). **d**, Immunoblotting analysis of *Endog* expression in ad.*Pgc1a*-infected cardiomyocytes (top), skeletal muscle of WT mice and transgenic mice expressing *Pgc1a* under the control of muscle creatine kinase (MCK-*Pgc1a*) (centre), and hearts of WT and cardiac-specific *Pgc1a*-deleted mice (*Pgc1a*^{AC/AC}) (bottom). **e**, *Endog* promoter activity, measured using a luciferase reporter, in HEK293 cells infected with ad.*Pgc1a* and/or ad.*Erra*. **f**, ERR- α ChIP-PCR of two regions of the *ENDOG* promoter. Red arrows denote primers, and ERRE specifies the location of a consensus ERR response element (1,304 bases upstream of the transcription start site). The experiment was repeated three times with similar results, and PCR products were quantified by quantitative PCR.

enriched for mitochondrial genes ($P = 2 \times 10^{-58}$) and oxidative metabolism processes ($P = 5 \times 10^{-38}$) (Fig. 3 and Supplementary Tables 2 and 3). Taken together, the high levels of *Endog* expression in metabolically active organs (Fig. 2a) and in brown fat (Supplementary Fig. 6), the unique co-expression of *ENDOG* with oxidative metabolism genes, and the link to AMPK signalling and ROS production

pointed to an unappreciated effect of *Endog* in physiological mitochondrial processes.

Peroxisome proliferator activated receptor- γ co-activator 1 α (PGC1 α) is widely recognized as a master regulator of mitochondrial function²⁴ and activates many target genes that are components of the *ENDOG*-associated network (Fig. 3b) through interaction with oestrogen-related receptor- α (ERR- α)⁹. Therefore, we tested whether PGC1 α also regulates *Endog*, and we observed robust PGC1 α -induced *Endog* transcript and *ENDOG* protein expression in cardiomyocytes *in vitro* (Fig. 3c, d). We confirmed the effects of varying *Pgc1a* expression on *ENDOG* protein expression *in vivo* using mice that overexpressed *Pgc1a* under the control of muscle creatine kinase (MCK-*Pgc1a*) and in mice in which *Pgc1a* had been deleted specifically in cardiomyocytes (*Pgc1a*^{AC/AC}) (Fig. 3d and Supplementary Methods). Luciferase studies revealed strong activation of the *Endog* promoter by PGC1 α and ERR- α together (Fig. 3e), and we confirmed direct binding of ERR- α to the *ENDOG* promoter by chromatin immunoprecipitation and PCR (ChIP-PCR) in a region containing an *ERRA* response element ($P < 0.001$) (Fig. 3f). These data show that *Endog* is a direct target of ERR- α and PGC1 α , master regulators of mitochondrial and heart function, further implicating *Endog* in mitochondrial and cardiac biology.

It was apparent that the effects of *Endog* loss of function on cardiac hypertrophy might be mediated through perturbations of mitochondrial physiology, which we therefore examined. Electron microscopy revealed lipid-like droplets associated with the mitochondria of *Endog*^{-/-} mice, and these droplets were more numerous and larger than those seen in control mice (Fig. 4a, b). Molecular studies revealed

a marked elevation of triglyceride levels in the hearts of *Endog*^{-/-} mice (Fig. 4c) that manifested as cardiomyocyte steatosis (Fig. 4a and Supplementary Fig. 7) but was not associated with variation in the expression levels of fatty acid metabolism or mitochondrial biogenesis genes (Supplementary Figs 8 and 9). Compared with their wild-type littermates, *Endog*^{-/-} mice had impaired mitochondrial respiration and increased ROS production (Fig. 4f, g).

To assess for mitochondrial depletion, we examined the ratio of mitochondrial DNA (mtDNA) to genomic DNA and the mitochondrial protein to tissue weight ratios, which were both diminished in the hearts of *Endog*^{-/-} mice (Fig. 4d, e) in the absence of mtDNA structural variation (Supplementary Fig. 10). This was an intriguing finding given the previously proposed roles for *Endog* in mtDNA synthesis, processing of polycistronic mtRNA and mitochondrial biogenesis^{25,26}, which had subsequently been discarded based primarily on experiments in *Endog*-deleted mice^{16,17}. We re-examined a role for *ENDOG* in mitochondrial biogenesis and demonstrated an increase in mitochondrial mass with chronic *ENDOG* expression in HEK293 cells ($P < 0.01$) and with acute *Endog* overexpression in a cardiomyocyte-derived cell line ($P < 0.001$) (Fig. 4h–k) in the absence of an effect on apoptotic or necrotic cell death (Supplementary Fig. 11). A role for *ENDOG* in mtDNA biology^{25,26} was supported further by ChIP-PCR experiments that showed direct binding of *ENDOG* throughout the mtDNA molecule (Fig. 4l), as previously demonstrated for mitochondrial transcription factor A (TFAM)²⁷, which is a crucial determinant of mtDNA synthesis and repair that when deleted causes eccentric cardiac hypertrophy and heart failure²⁸.

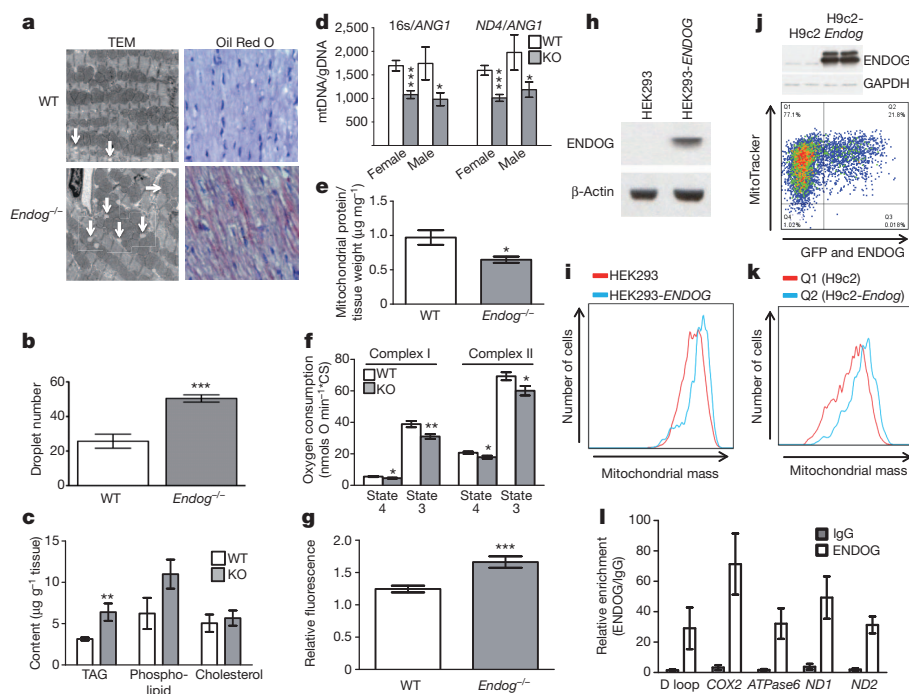


Figure 4 | *Endog* regulates mitochondrial function and cardiac lipid metabolism. **a**, Transmission electron micrographs (TEM) and oil red O stained micrographs (high resolution in Supplementary Fig. 7) of left ventricular sections from WT and *Endog*^{-/-} mice. Vertical arrows indicate lipid droplets, and the horizontal arrow indicates a mitochondrion. **b**, Quantification of the number of mitochondria-associated droplets in WT and *Endog*^{-/-} mice. **c**, Quantification of cardiac triglyceride (TAG), phospholipid and cholesterol content in the hearts of WT and *Endog*^{-/-} (KO) mice ($n = 5$). **d**, Ratio of mtDNA to genomic DNA (gDNA) in the hearts of WT and *Endog*^{-/-} (KO) mice. **e**, Quantification of mitochondrial protein content in WT and *Endog*^{-/-} mice ($n = 5$). **f**, State 3 and state 4 oxygen consumption in the presence of substrates for complex I or complex II of the electron transport chain in cardiac mitochondria isolated from WT ($n = 6$) or *Endog*^{-/-} (KO) ($n = 5$) mice. *CS, corrected for citrate synthase activity. **g**, Relative

fluorescence-based measurement of ROS production by mitochondria isolated from WT ($n = 6$) or *Endog*^{-/-} ($n = 5$) mice. **h–k**, Representative flow cytometric analysis of mitochondrial mass in HEK293 and H9c2 cells overexpressing human or rat *ENDOG*, respectively ($n = 4$). **h**, Stable expression of *ENDOG* in HEK293 cells (HEK293-*ENDOG*). **i**, Flow cytometric analysis of HEK293 and HEK293-*ENDOG* cells stained with MitoTracker. **j**, Adenovirus (ad)-mediated expression of GFP and rat *ENDOG* in cardiomyocytes (top), and flow cytometric analysis of ad-*Endog*-infected cells (Q2) and uninfected control cells (Q1). Colours denote the event density: from highest to lowest density, red, orange, green then blue. **k**, Number of events plotted against mitochondrial mass for ad-*Endog*-infected (Q2) and control (Q1) H9c2 cells. **l**, Quantitative PCR of mtDNA–protein complexes after ChIP of mitochondrial chromatin using anti-*ENDOG* antibody or IgG. All data are presented as mean \pm s.e.m. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Mitochondria are essential for oxidative metabolism, and mitochondrial dysfunction and/or depletion in the heart causes maladaptive cardiac hypertrophy and cardiac dysfunction associated with increased amounts of ROS and lipotoxicity^{4,5,28,29}. Here we identified *Endog* loss of function as a primary determinant of maladaptive cardiac hypertrophy that is associated with mitochondrial dysfunction and depletion and with marked cardiomyocyte steatosis. The mechanism underlying cardiac hypertrophy that results from impaired mitochondrial function is not limited to a single pathway, but we demonstrated a conserved increase in ROS, which are established hypertrophic stimuli^{21,22}, in *Endog* loss-of-function models. Our studies resolve some of the uncertainty about the non-apoptotic function of *Endog*^{15–17} and reveal its importance in mitochondrial biology, which has intriguing parallels with the dual roles of apoptosis-inducing factor³⁰. We propose that *ENDO*G, which we show binds to mtDNA, modulates mtDNA synthesis, maintenance and/or transcription, which is consistent with previous hypotheses^{25,26}. Therapeutic targeting of the *PGC1A*–*ERRA* axis has been proposed to improve mitochondrial function in cardiac failure¹¹, and our studies suggest that regulation of *Endog* is an important component of this process. We conclude that *Endog* is a novel determinant of maladaptive cardiac hypertrophy with previously unappreciated mitochondrial functions.

METHODS SUMMARY

Linkage mapping was carried out using microsatellite genotypes in the BN × SHR F₂ population. *Ex vivo* heart weight analysis was performed in the congenic rat strains, which were characterized using *in vivo* blood pressure telemetry. Comparative haplotype analysis was performed using single nucleotide polymorphism data (Rat Genome Database; <http://rgd.mcw.edu/>) for all strains used in the QTL mapping studies. Microarray-based expression analysis was conducted as described previously^{6,7}. Cell size and hypertrophy biomarker expression were measured in cardiomyocytes after lentivirus-mediated *Endog* knockdown. Heart weight, hypertrophic biomarker expression and cardiomyocyte size were measured in *Endog*^{−/−} mice at baseline and after angiotensin-II-induced hypertrophy. Triglyceride abundance, mitochondrial mass and respiratory activity were measured in *Endog*^{−/−} mice as described in the Supplementary Information. Weighted gene co-expression network analysis (WGCNA)²³ was applied to the largest publicly available human heart transcriptome data set. Regulation of *Endog* by *PGC1α* was investigated in neonatal cardiomyocytes infected with adenovirus expressing *Pgc1a*, in MCK-*Pgc1a* skeletal muscle and in *Pgc1a*^{ΔC/ΔC} heart samples. The association of *ERR-α* with the *ENDO*G promoter and the *ENDO*G–mtDNA interaction were determined using ChIP. The histological analysis and electron microscopy of *Endog*^{−/−} hearts was carried out to study mitochondrial structure and abundance, as well as lipid deposition. Genomic DNA and mtDNA copy number were assessed by quantitative PCR. The mitochondrial abundance in cells was studied by flow cytometry. Full methods are provided in the Supplementary Methods.

Received 6 March; accepted 17 August 2011.

- Post, W. S., Larson, M. G., Myers, R. H., Galderisi, M. & Levy, D. Heritability of left ventricular mass: the Framingham Heart Study. *Hypertension* **30**, 1025–1028 (1997).
- Lorell, B. H. & Carabello, B. A. Left ventricular hypertrophy: pathogenesis, detection, and prognosis. *Circulation* **102**, 470–479 (2000).
- Vasan, R. S. *et al.* Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *J. Am. Med. Assoc.* **302**, 168–178 (2009).
- McGavock, J. M., Victor, R. G., Unger, R. H. & Szczepaniak, L. S. Adiposity of the heart, revisited. *Ann. Intern. Med.* **144**, 517–524 (2006).
- Wong, C. & Marwick, T. H. Obesity cardiomyopathy: pathogenesis and pathophysiology. *Nature Clin. Pract. Cardiovasc. Med.* **4**, 436–443 (2007).
- Heinig, M. *et al.* A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature* **467**, 460–464 (2010).
- Petretto, E. *et al.* Integrated genomic approaches implicate osteoglycin (*Ogn*) in the regulation of left ventricular mass. *Nature Genet.* **40**, 546–552 (2008).
- Li, L. Y., Luo, X. & Wang, X. Endonuclease G is an apoptotic DNase when released from mitochondria. *Nature* **412**, 95–99 (2001).
- Dufour, C. R. *et al.* Genome-wide orchestration of cardiac functions by the orphan nuclear receptors *ERRα* and *γ*. *Cell Metab.* **5**, 345–356 (2007).
- Wu, Z. *et al.* Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator *PGC-1*. *Cell* **98**, 115–124 (1999).

- Finck, B. N. & Kelly, D. P. *PGC-1* coactivators: inducible regulators of energy metabolism in health and disease. *J. Clin. Invest.* **116**, 615–622 (2006).
- Hill, J. A. & Olson, E. N. Cardiac plasticity. *N. Engl. J. Med.* **358**, 1370–1380 (2008).
- Inomata, H. *et al.* Identification of quantitative trait loci for cardiac hypertrophy in two different strains of the spontaneously hypertensive rat. *Hypertens. Res.* **28**, 273–281 (2005).
- Siegel, A. K. *et al.* Genetic loci contribute to the progression of vascular and cardiac hypertrophy in salt-sensitive spontaneous hypertension. *Arterioscler. Thromb. Vasc. Biol.* **23**, 1211–1217 (2003).
- Büttner, S. *et al.* Endonuclease G regulates budding yeast life and death. *Mol. Cell* **25**, 233–246 (2007).
- David, K. K., Sasaki, M., Yu, S. W., Dawson, T. M. & Dawson, V. L. EndoG is dispensable in embryogenesis and apoptosis. *Cell Death Differ.* **13**, 1147–1155 (2006).
- Irvine, R. A. *et al.* Generation and characterization of endonuclease G null mice. *Mol. Cell Biol.* **25**, 294–302 (2005).
- Temme, C. *et al.* The *Drosophila melanogaster* gene *cg4930* encodes a high affinity inhibitor for endonuclease G. *J. Biol. Chem.* **284**, 8337–8348 (2009).
- Bahi, N. *et al.* Switch from caspase-dependent to caspase-independent death during heart development: essential role of endonuclease G in ischemia-induced DNA processing of differentiated cardiomyocytes. *J. Biol. Chem.* **281**, 22943–22952 (2006).
- Arad, M. *et al.* Constitutively active AMP kinase mutations cause glycogen storage disease mimicking hypertrophic cardiomyopathy. *J. Clin. Invest.* **109**, 357–362 (2002).
- Dai, D. F. *et al.* Mitochondrial oxidative stress mediates angiotensin II-induced cardiac hypertrophy and Gαq overexpression-induced heart failure. *Circ. Res.* **108**, 837–846 (2011).
- Seddon, M., Looi, Y. H. & Shah, A. M. Oxidative stress and redox signalling in cardiac hypertrophy and heart failure. *Heart* **93**, 903–907 (2007).
- Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17 (2005).
- Spiegelman, B. M. Transcriptional control of mitochondrial energy metabolism through the *PGC1* coactivators. *Novartis Found. Symp.* **287**, 60–69 (2007).
- Cote, J. & Ruiz-Carrillo, A. Primers for mitochondrial DNA replication generated by endonuclease G. *Science* **261**, 765–769 (1993).
- Tiranti, V. *et al.* Chromosomal localization of mitochondrial transcription factor A (TCF6), single-stranded DNA-binding protein (SSBP), and endonuclease G (ENDO)G, three human housekeeping genes involved in mitochondrial biogenesis. *Genomics* **25**, 559–564 (1995).
- Rothfuss, O. *et al.* Parkin protects mitochondrial genome integrity and supports mitochondrial DNA repair. *Hum. Mol. Genet.* **18**, 3832–3850 (2009).
- Wang, J. *et al.* Dilated cardiomyopathy and atrioventricular conduction blocks induced by heart-specific inactivation of mitochondrial DNA gene expression. *Nature Genet.* **21**, 133–137 (1999).
- Lewis, W. *et al.* Decreased mtDNA, oxidative stress, cardiomyopathy, and death from transgenic cardiac targeted human mutant polymerase *γ*. *Lab. Invest.* **87**, 326–335 (2007).
- Vahsen, N. *et al.* AIF deficiency compromises oxidative phosphorylation. *EMBO J.* **23**, 4679–4689 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge funding from the Medical Research Council (UK), the National Institute for Health Research (UK), the Royal Brompton and Harefield Cardiovascular Biomedical Research Unit, the Imperial College Healthcare Biomedical Research Centre, the British Heart Foundation, Fondation Leducq, the Wellcome Trust, the Grant Agency of the Czech Republic (301/08/0166), the Ministry of Education of the Czech Republic (1M0520), the Ministerio de Ciencia e Innovación (Spain; PTQ-08-03-07880, SAF2008-02271, SAF2008-03067 and SAF2010-19125), the Agència de Gestió d'Ajuts Universitaris i Recerca (Spain; 2009-SGR-346), the Fondo de Investigaciones Sanitarias (Spain; PS09/02034, PS09/01602 and PS09/01591), the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. HEALTH-F4-2010-241504 (EURATRANS), and the German National Genome Research Network (NGFN-Plus) Heart Failure. We thank M. R. Lieber for providing the *Endog* deleted mice and E. Wahle for providing the CG4930 expression plasmid. We thank the National BioResource Project for the Rat (<http://www.anim.med.kyoto-u.ac.jp/NBR/>) for providing rat strains.

Author Contributions C.M.-R., J.Y., X.-M.S., A.S., J.Z., A.B., R.B., D.H., H.L., G.C.R., R.M. and E.G.-A. performed the laboratory-based experiments. R.A., P.M., M.M., V.Z., F.P., M.C., M.R.-M. and F.K. performed the physiology experiments. N.H., H.J., L.E.F., P.J.R.B. and T.S. provided gene expression and physiology data. J.W., L.B. and E.P. performed genetic mapping and network studies. X.C., J.X.C., Z.A., M.P. and D.C.-D. supervised the data analysis and contributed to the experimental design. S.A.C. and D.S. planned the experiments. S.A.C. wrote the manuscript with input and discussion from all of the co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.A.C. (stuart.cook@imperial.ac.uk) or D.S. (daniel.sanchis@cmb.udl.cat).

Control of flowering and storage organ formation in potato by FLOWERING LOCUS T

Cristina Navarro^{1*}, José A. Abelenda^{1*}, Eduard Cruz-Oró¹, Carlos A. Cuéllar¹, Shojiro Tamaki², Javier Silva¹, Ko Shimamoto² & Salomé Prat¹

Seasonal fluctuations in day length regulate important aspects of plant development such as the flowering transition or, in potato (*Solanum tuberosum*), the formation of tubers. Day length is sensed by the leaves, which produce a mobile signal transported to the shoot apex or underground stems to induce a flowering transition or, respectively, a tuberization transition. Work in *Arabidopsis*, tomato and rice (*Oryza sativa*) identified the mobile FLOWERING LOCUS T (FT) protein as a main component of the long-range 'florigen', or flowering hormone, signal^{1–3}. Here we show that expression of the *Hd3a* gene, the FT orthologue in rice, induces strict short-day potato types⁴ to tuberize in long days. Tuber induction is graft transmissible and the *Hd3a*–GFP protein is detected in the stolons of grafted plants, transport of the fusion protein thus correlating with tuber formation. We provide evidence showing that the potato floral and tuberization transitions are controlled by two different FT-like paralogues (StSP3D and StSP6A) that respond to independent environmental cues, and show that an autorelay mechanism involving CONSTANS modulates expression of the tuberization-control StSP6A gene.

Potato, the third largest global food crop after wheat and rice, is cultivated for its underground storage stems or tubers, which are rich in starch and other nutrients. Short days and cool temperatures promote tuber formation, ensuring that differentiation of these vegetative propagation organs precedes winter. Whereas cultivated potatoes derive from Chilean landraces more adapted to long-day conditions, Andean types such as the *S. tuberosum* group Andigena (*Solanum tuberosum andigena*) tuberize only in short days⁵. These plants require night periods longer than a critical length, as a pulse of light during the night (a 'night break') represses tuberization, as seen in strict short-day flowering plants^{6,7}.

Day length is sensed by expanded leaves, which synthesize a mobile signal or 'tuberigen' that is transported to the underground stems to induce tuber formation⁴. This long-distance signal shares several features with the mobile florigen, with different pieces of evidence suggesting that related photoperiodic pathways may control synthesis of both signals⁸. In *Arabidopsis thaliana*, activation of the FT gene by CONSTANS (CO) mediates floral transition in long days^{9,10}. Transport of the FT protein from the leaf to the shoot apical meristem has been demonstrated in diverse ways^{11–13}, it being broadly accepted that this phloem mobile protein functions as the florigen. Closely related genes also mediate control of flowering in rice, although in this short-day plant the CO orthologue *Heading date 1*¹⁴ (*Hd1*) activates expression of the FT-like *Hd3a* gene in short days but represses its expression under long-day conditions^{15,16}.

A CO-dependent pathway is also thought to mediate short-day tuberization, as expression of *Arabidopsis CO* in Andigena plants delays tuber formation in short days¹⁷. High light irradiance, however, induces Andigena potatoes to flower in long days, although these conditions are restrictive for tuberization⁸. This long-day flowering

response seemed to argue against a role for FT in tuberization, implicating an additional CO target as the mobile tuberigen. Here we show that ectopic expression of the rice *Hd3a* gene induces Andigena plants to flower and tuberize under non-inductive long days, demonstrating the potential of FT to act as the mobile tuberigen. We show, in addition, that flowering and short-day tuberization responses are regulated by two members of the potato FT-like gene family that respond to different environmental cues.

To assess whether FT has a role in tuberization, we transformed Andigena plants with the *rolC::Hd3a*–GFP construct (GFP, green fluorescent protein), which in rice promotes floral transition in long days¹⁸. Lines expressing this construct were induced to flower (Fig. 1b, c) and were able to tuberize in non-inductive long days (Fig. 1a). When grafted to wild-type plants, these lines induced the wild-type controls to tuberize in long days, independently of whether they were used as donors (*Hd3a* onto wild type) or as stocks (wild type onto *Hd3a*), whereas none of the control grafts (wild type onto wild type; Fig. 1d) tuberized in non-inductive long days. We detected the *Hd3a*–GFP protein but not its transcript in the stolons of grafted wild-type stocks (Supplementary Fig. 1a, c), demonstrating that the protein but not the RNA can move across the graft junction and function as a powerful tuberization inducer.

Studies in tomato identified six members of the *SELF-PRUNING* (SP) gene family¹⁹ and sequencing of two diploid potato (Tuberosum RH89-039-16 and Phureja DM1-3 516 R44) genomes recently led to the identification of three additional FT and TFL family members²⁰. Phylogeny of these genes grouped the StSP6A, StSP5G, StSP5G-like and StSP3D homologues into the same clade as the *Arabidopsis FT*, tomato *SINGLE-FLOWER TRUSS* (SFT) and rice *Hd3a* genes (Supplementary Fig. 2). Quantitative PCR with reverse transcription (RT–PCR) revealed that these transcripts are expressed in leaves or stolons, whereas transcripts for *SP/TFL1* and *MFT* members are more ubiquitously distributed (Fig. 1e). Interestingly, StSP6A gene expression strongly correlates with tuberization, high levels of expression being observed in leaves and stolons of short-day-induced plants and in antisense phytochrome B lines, with constitutive tuberization²¹ (Fig. 1f and Supplementary Information, section 1). This expression profile suggests that this gene is involved in tuberization control, StSP6A overexpression (StSP6Aox) lines being able to tuberize under non-inductive long days (Fig. 2a, b) and induced to flower, although their flowering phenotype is less severe than in *Hd3a* lines (Supplementary Fig. 3a, b and Supplementary Information, section 2). StSP6A silencing, in turn, strongly delays tuber formation in short days (Fig. 2c, d), pointing to an essential role for this FT-like protein in tuberization promotion. StSP6A expression analyses in commercial cultivars with early (Jaerla), late (Baraka) and intermediate (Kennebec) tuberization periods, in addition, show that accumulation of this transcript in leaves correlates with the tuberization time of these cultivars (Supplementary Fig. 6), indicating that this FT paralogue is involved in tuberization control even in non-photoperiodic cultivars.

¹Departamento de Genética Molecular de Plantas, Centro Nacional de Biotecnología-CSIC. Calle Darwin 3, 28049 Madrid, Spain. ²Laboratory of Plant Molecular Genetics, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma 630-0101, Japan.

*These authors contributed equally to this work.

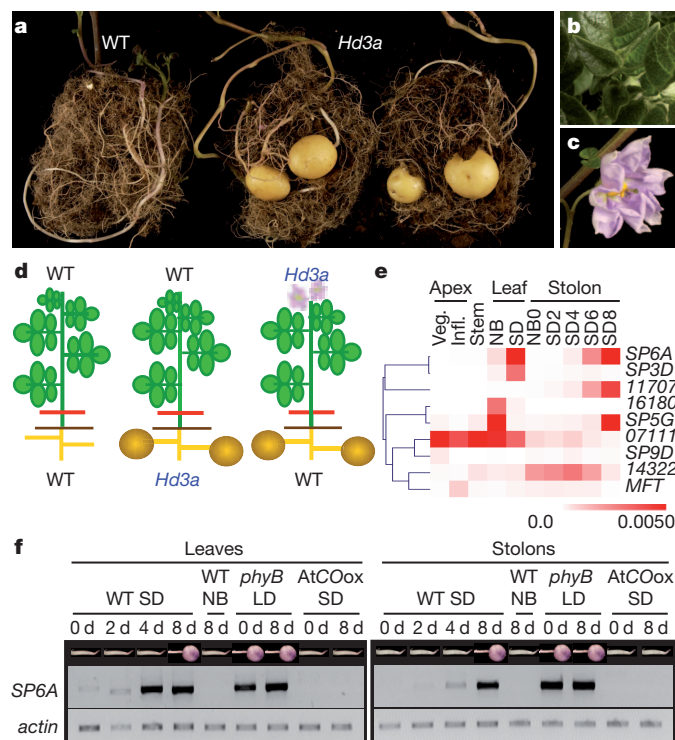


Figure 1 | Phenotype of *Andigena rolC::Hd3a*-GFP lines and expression profiles of the potato *FT*- and *TFL1*-like genes. **a**, *Hd3a* lines (centre and right) tuberize under long-day conditions. Left, wild-type (WT) control. **b**, **c**, Early-flowering *Hd3a* plants (**c**) relative to wild type (**b**). **d**, Tuber induction in long days of wild-type plants grafted with *Hd3a* donors (right) or stocks (centre). Wild-type control grafts do not tuberize (left). Red lines indicate the graft junction. **e**, Relative levels of expression of the potato *FT*- and *TFL1*-like genes. 11707, PGSC0003DMG400011707; 16180, PGSC0003DMG400016180; 07111, PGSC0003DMG400007111; 14322, PGSC0003DMG400014322; Veg., vegetative apex; Inf., inflorescence apex; NB, night break; SD, short day. **f**, Semiquantitative RT-PCR analysis of *StSP6A* expression in leaves and stolons of plants with different tuberization states (indicated on top). The pictures show non-induced stolons or tubers. LD, long day.

Day-neutral tomato flowering is regulated by the *FT*-homologue *SFT/SP3D* gene². This gene has been reported to be regulated independently of CO and day length^{22,23}, its potato orthologue being likely to have a role in light-irradiance-dependent flowering. *StSP3D*-silenced lines were actually found to be late flowering (Fig. 2e–h), although under short-day conditions they tuberized at the same time as untransformed controls (Supplementary Fig. 7 and Supplementary Information, section 2), indicating that this gene is important in flowering but not in the tuberization transition. Remarkably, in modern tomato cultivars *SP6A* was reported to have an extra nucleotide (T at position 421) that leads to a premature stop codon, implying a non-functional role for this gene¹⁹. *StSP3D*, in turn, retains a weak short-day activation response in *Andigena*, although transcript levels are lower than those of *StSP6A* (Fig. 1e). Together, these observations suggest that members of the *FT*-like family in Solanaceae diversified such that *SFT/SP3D* ceased to be regulated by CO^{22,23} and became responsive to other environmental cues, such as to be important in day-neutral flowering control. *SP6A* function was lost in modern tomato cultivars but in potato plays a major part in tuberization, as day-length-dependent activation of this gene mediates strict short-day tuberization of *Andigena* species. Thus, the roles of these two *FT* paralogues in flowering and tuberization control have evolved, at least in part, through changes in their expression profiles, both genes encoding for functionally similar proteins, as indicated by the finding that *StSP6A* expression in *Arabidopsis* rescues the late-flowering phenotype of *co-1* and *ft-1* mutants (Supplementary Fig. 8).

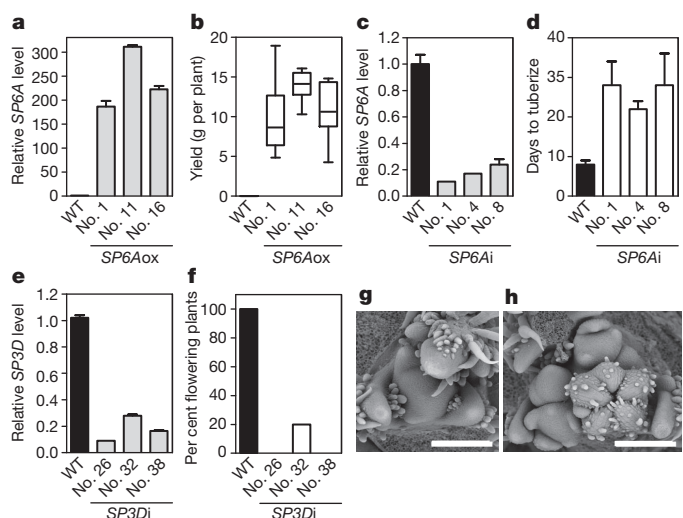


Figure 2 | Function of the *StSP6A* and *StSP3D* genes in short-day-dependent tuberization and day-neutral flowering control. **a**, **b**, *StSP6A* expression in leaves of *StSP6Aox* lines (**a**) and tuber induction in these lines (**b**) in long days. **c**, **d**, *StSP6A* silencing in leaves of *StSP6A* RNAi (*SP6Ai*) lines (**c**) correlates with late tuberization in short days (**d**). **e**, Relative *StSP3D* expression in *StSP3D* RNAi (*SP3Di*) lines. Error bars, s.d. ($n > 3$). **f**, Late flowering of *StSP3D*-silenced lines. **g**, **h**, Scanning electron microscopy images of the *StSP3D* RNAi apices: vegetative apex in L26 (**g**); apex in inflorescence transition in L32 (**h**). Scale bars, 200 μ m.

Interestingly, two other *FT*-clade members, *StSP5G* and *StSP5G*-like (Supplementary Table 2), are expressed under non-inductive long-day conditions (Fig. 1e). This expression profile suggests an antagonistic function for *StSP6A* and these two *FT* paralogues (Supplementary Information, section 3), as recently reported for the sugar beet (*Beta vulgaris*) *BvFT1* and *BvFT2* genes²⁴. Further studies will be required to assess whether these *FT*-clade members have an inhibitory effect on tuberization.

StSP6A is expressed not only in leaves but also in stolons of tuberizing plants (Fig. 1f). Expression in these organs is delayed with respect to the leaves, suggesting an autoregulatory loop for the transported protein. In support of this relay mechanism, increased levels of expression of the endogenous *StSP6A* transcript are observed in *Hd3a* lines (Supplementary Fig. 10) and wild-type stocks grafted to these plants (Supplementary Fig. 10a). Thus, in contrast to the *Arabidopsis FT* or rice *Hd3a* gene, *StSP6A* is regulated by a relay mechanism that sustains synthesis of the inducing signal in stolons. In this regard, a local balance between the SP floral repressor and the tomato *SFT* florigen signal has been shown to contribute to the differential flowering response of primary and secondary shoot meristems²⁵. Hence, it is possible that having more SP in stolons confers a reduced sensitivity to FT, which may explain why floral transition is activated without tuber formation, whereas *Hd3a* overexpression activates both developmental processes.

To rule out that tuberization of grafted stock plants is mostly mediated by this relay mechanism, we grafted *Hd3a* grafts onto *StSP6A* RNA interference (RNAi) stocks to test whether inhibition of *StSP6A* expression blocks *Hd3a* signalling. As shown in Fig. 3a, activation of the endogenous *StSP6A* gene is strongly reduced in *StSP6A* RNAi compared with the wild-type stocks. Tuberization onset is delayed and tuber yield reduced in RNAi lines relative to the *Hd3a*/wild-type grafted controls (Fig. 3b), owing to impaired signal amplification, but this does not preclude tuberization of the RNAi stocks, highlighting induction by the *Hd3a* protein.

Finally, we tested whether this regulatory relay requires CO function, by grafting *StSP6Aox* plants into *StCOox* stocks and analysing endogenous *StSP6A* gene expression (Fig. 3c–e). Like the rice Hd1 protein^{15,16}, *StCO* represses *StSP6A* gene expression in long days, and

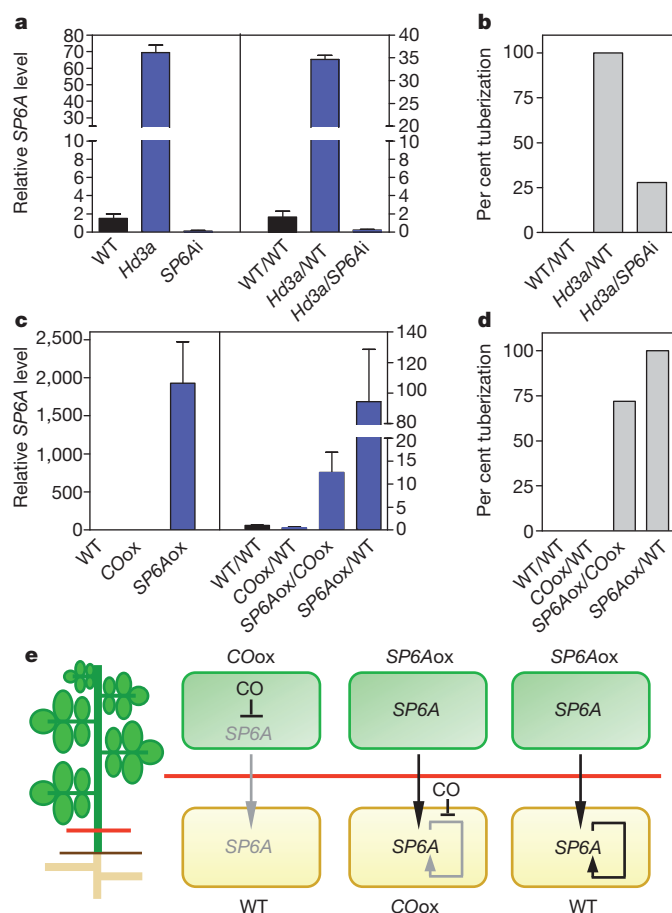


Figure 3 | Autoregulatory loop for StSP6A expression. **a**, Levels of StSP6A transcript in transgenic (left) and grafted (right) plants. *Hd3a* scions were grafted onto wild-type or StSP6A RNAi (SP6Ai) lines. **b**, Tuber induction in grafted plants, scored when 100% of the *Hd3a*/wild-type grafts were tuberizing. **c**, Relative expression of the endogenous StSP6A transcript in COox and SP6Aox plants (left), and in grafts with wild-type and COox stocks (right). **d**, Tuber induction of the grafted plants grown in non-inductive long days. Error bars, s.d. of two experimental replicates ($n > 8$). See Supplementary Fig. 12 for each replicate result. **e**, StSP6A autoregulation.

transfer to short-day conditions relieves this repression (Supplementary Information, section 4). In line with this function, activation of the StSP6A gene is largely repressed in stolons of the StSP6Aox/StCOox grafted plants (Fig. 3c), relative to StSP6Aox/wild-type grafts used as controls, which implies that StCO is involved in the autoregulatory loop that drives StSP6A expression in stolons. Moreover, StSP6A accumulates only to basal levels in the stolons of StCOox/wild-type and wild-type/wild-type grafts used as negative controls (StSP6A is not expressed in long days in StCOox or wild-type scions), which implies that expression in these organs requires the mobile protein produced in the leaves.

Our data provide the molecular basis for a long-standing physiological observation, namely that flowering tobacco plants grafted onto non-induced potato stocks induce tuberization of the stock plants, irrespective of the photoperiodic requirement of the donor plant²⁶. An additional issue is how flowering and tuberization transitions are differentially triggered in response to the mobile FT signal. In *Arabidopsis*, FT interacts with FD to activate expression of floral meristem identity genes^{27,28}. We observed that *Hd3a* stolons initiate floral buds from the apical meristem (Fig. 4a–c) at the same time as they differentiate tubers from the subapical region. Using transgenic lines expressing the StSP6A protein under control of an ethanol-inducible promoter, tuber-specific transcripts were observed within 4 h of induction in the stolons (Fig. 4d, e and Supplementary Information,

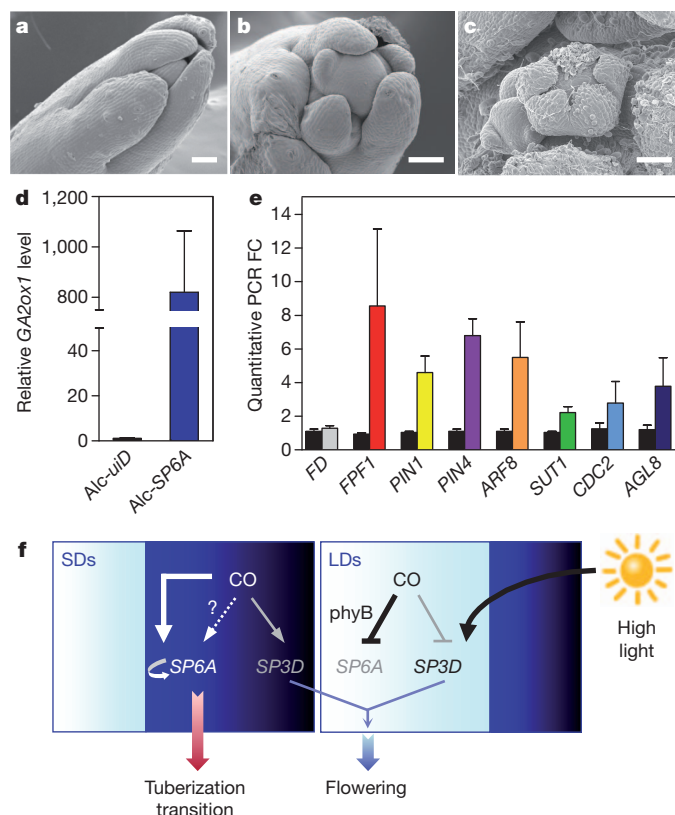


Figure 4 | Floral phenotype of *Hd3a* stolons and proposed model for flower and tuber induction. **a–c**, Flower induction in *Hd3a* stolons. Scanning electron microscope images of wild-type (**a**) and *Hd3a* (**b**, **c**) stolon apical meristems. Scale bars, 100 μ m. **d**, Local StSP6A induction activates StGA2ox1 gene expression in the stolon. **e**, Activation of other genes reported to be induced early during tuber development. Colour bars represent fold change (FC) in gene expression in Alc-StSP6A stolons relative to Alc-uid controls (black bars). Alc, ethanol-inducible promoter; error bars, s.d. of two experimental replicates ($n > 10$). **f**, Model for regulation of flowering and tuberization transitions. See Supplementary Information, section 6, for a detailed explanation.

section 5). This very rapid response precludes transport of any mobile signal from the leaves and thus supports the notion that the StSP6A protein is the mobile tuberigen transported to below-ground organs. Thus, it is possible that StSP6A interacts in stolon subapical cells with an as yet unknown transcriptional regulator, or that FD–StSP6A differentially activates a set of target genes specific to these cells, such as StGA2ox1 (ref. 29; Fig. 4d). A challenge for the future is to identify such stolon-specific StSP6A target genes and/or StSP6A-interacting partners, to establish how formation of these storage organs is regulated.

Identification of FT as a switch for tuberization supports the notion that FT function extends beyond flowering induction (Fig. 4f). Other studies have implicated FT in seasonal control of growth cessation in poplar trees³⁰ and meristem growth termination in tomato²⁵. Thus, FT is emerging as a key mobile signal controlling not only flowering but also a number of other meristem-associated transitions.

METHODS SUMMARY

Plant materials and growth conditions. Andigena 7540 wild-type plants and antisense phytochrome B²¹ and AtCOox (ref. 17) lines were grown in a greenhouse under long-day conditions. Growth chambers were used for controlled short-day (8 h light, 16 h dark) and short-day/night-break (short day plus a 30-min pulse of light in the middle of the night) treatments. Transgenic Andigena plants were generated by *Agrobacterium*-mediated transformation of leaf explants. Plants were grafted as previously described¹⁷ and cultivated under long-day conditions to analyse their tuberization response.

Plasmid constructs. StSP6A was cloned into the pBinAR binary vector to generate StSP6A-overexpressing lines. StSP6A, StSP3D and StCO RNAi constructs were

generated by inserting the 3' non-conserved regions of these genes in opposite orientations into the pBIN19RNAi vector. The pGWB2 vector was used for StCO overexpression. StSP6A ethanol-inducible lines were generated by transformation with the pBinSRNA-GW plasmid containing the StSP6A coding region. Primer sets used for these constructs are listed in Supplementary Table 3.

Real-time RT-PCR analyses. First-strand complementary DNA was synthesized from 2 µg total RNA and 1 µl of the reaction used for real-time gene expression analysis with the SYBR Green PCR master mix (Applied Biosystems). The *actin 8* gene was used for normalization. Identical procedures were used for semiquantitative RT-PCR, except that amplification was conducted in a Peltier thermal cycler (PTC-200, MJ Research). For sets of specific primers and product lengths, see Supplementary Table 3.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 December 2010; accepted 8 August 2011.

Published online 25 September 2011.

- Turck, F., Fornara, F. & Coupland, G. Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. *Annu. Rev. Plant Biol.* **59**, 573–594 (2008).
- Lifschitz, E. *et al.* The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. *Proc. Natl Acad. Sci. USA* **103**, 6398–6403 (2006).
- Zeevaert, J. A. Leaf-produced floral signals. *Curr. Opin. Plant Biol.* **11**, 541–547 (2008).
- Jackson, S. D. Multiple signalling pathways control tuber induction in potato. *Plant Physiol.* **119**, 1–8 (1999).
- Ewing, E. E. & Struik, P. C. Tuber formation in potato: induction, initiation and growth. *Hortic. Rev. (Am. Soc. Hortic. Sci.)* **14**, 89–197 (1992).
- Ishikawa, R. *et al.* Suppression of the floral activator Hd3a is the principal cause of the night break effect in rice. *Plant Cell* **17**, 3326–3336 (2005).
- Lagercrantz, U. At the end of the day: a common molecular mechanism for photoperiod responses in plants? *J. Exp. Bot.* **60**, 2501–2515 (2009).
- Rodríguez-Falcón, M., Bou, J. & Prat, S. Seasonal control of tuberization in potato: conserved elements with the flowering response. *Annu. Rev. Plant Biol.* **57**, 151–180 (2006).
- Samach, A. *et al.* Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. *Science* **288**, 1613–1616 (2000).
- An, H. *et al.* CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of *Arabidopsis*. *Development* **131**, 3615–3626 (2004).
- Corbesier, L. *et al.* FT protein movement contributes to long-distance signaling in floral induction of *Arabidopsis*. *Science* **316**, 1030–1033 (2007).
- Jaeger, K. E. & Wigge, P. A. FT protein acts as a long-range signal in *Arabidopsis*. *Curr. Biol.* **17**, 1050–1054 (2007).
- Mathieu, J., Warthmann, N., Kuttner, F. & Schmid, M. Export of FT protein from phloem companion cells is sufficient for floral induction in *Arabidopsis*. *Curr. Biol.* **17**, 1055–1060 (2007).
- Yano, M. *et al.* Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS. *Plant Cell* **12**, 2473–2484 (2000).
- Kojima, S. *et al.* Hd3a, a rice ortholog of the *Arabidopsis* FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. *Plant Cell Physiol.* **43**, 1096–1105 (2002).
- Hayama, R., Yokoi, S., Tamaki, S., Yano, M. & Shimamoto, K. Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* **422**, 719–722 (2003).
- Martinez-Garcia, J. F., Virgos-Soler, A. & Prat, S. Control of photoperiod-regulated tuberization in potato by the *Arabidopsis* flowering-time gene CONSTANS. *Proc. Natl Acad. Sci. USA* **99**, 15211–15216 (2002).
- Tamaki, S., Matsuo, S., Wong, H. L., Yokoi, S. & Shimamoto, K. Hd3a protein is a mobile flowering signal in rice. *Science* **316**, 1033–1036 (2007).
- Carmel-Goren, L., Liu, Y. S., Lifschitz, E. & Zamir, D. The SELF-PRUNING gene family in tomato. *Plant Mol. Biol.* **52**, 1215–1222 (2003).
- The Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature* **475**, 189–195 (2011).
- Jackson, S. D., Heyer, A., Dietze, J. & Prat, S. Phytochrome B mediates the photoperiodic control of tuber formation in potato. *Plant J.* **9**, 159–166 (1996).
- Ben-Naim, O. *et al.* The CCAAT binding factor can mediate interactions between CONSTANS-like proteins and DNA. *Plant J.* **46**, 462–476 (2006).
- Lifschitz, E. & Eshed, Y. Universal florigenic signals triggered by FT homologues regulate growth and flowering cycles in perennial day-neutral tomato. *J. Exp. Bot.* **57**, 3405–3414 (2006).
- Pin, P. A. *et al.* An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. *Science* **330**, 1397–1400 (2010).
- Shalit, A. *et al.* The flowering hormone florigen functions as a general systemic regulator of growth and termination. *Proc. Natl Acad. Sci. USA* **106**, 8392–8397 (2009).
- Chailakhyan, M. K., Yanina, L. I., Davedzhyan, A. G. & Lotova, G. N. Photoperiodism and tuber formation in grafting of tobacco onto potato. *Dokl. Akad. Nauk SSSR* **257**, 1276–1280 (1981).
- Abe, M. *et al.* FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* **309**, 1052–1056 (2005).
- Wigge, P. A. *et al.* Integration of spatial and temporal information during floral induction in *Arabidopsis*. *Science* **309**, 1056–1059 (2005).
- Kloosterman, B. *et al.* StGA2ox1 is induced prior to stolon swelling and controls GA levels during potato tuber development. *Plant J.* **52**, 362–373 (2007).
- Bohlenius, H. *et al.* CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science* **312**, 1040–1043 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Paz-Ares, G. Bryan and C. Bachem for their comments on the manuscript. We also thank S. Yokoi for his help in this work. This research was supported by grants from the Spanish MCYT and the EU-SOL European Union Integrated Project. Support by the JSPS and CSIC under the Japan–Spain research cooperative programme and Grants-in-Aid for Scientific Research on Priority Areas of the MEXT of Japan are also acknowledged. C.N. was recipient of an I3P postdoctoral contract from the CSIC.

Author Contributions C.N., J.A.A., E.C.-O., C.A.C., J.S. and S.P. performed experiments and analysed the results. S.T. and K.S. provided the rice Hd3a construct and performed the Hd3a protein mobility studies. S.P. designed the experiments and wrote the manuscript.

Author Information Microarray data on StSP6Aox, StSP6A RNAi and stolon gene expression during photoperiod induction have been deposited in the Gene Expression Omnibus under the accession codes GSE31178 and GSE31336. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.P. (sprat@cnb.csic.es).

METHODS

Plant materials and growth conditions. Andigena 7540 wild-type lines and the antisense phytochrome B²¹ (*phyB*) and AtCO overexpressor¹⁷ (AtCOox) lines were grown in the greenhouse under long-day conditions. Growth chambers were used for controlled short-day (8 h light, 16 h dark), short-day/night-break (short day plus a 30-min pulse of light given 8 h after the beginning of the dark period) and long-day (16 h light/8 h dark) treatments.

Transgenic Andigena plants carrying the *rolC::Hd3a-GFP*¹⁸, *StSP6A* overexpression, *StSP6A* RNAi, *StSP3D* RNAi, *StCOox* and *StCO* RNAi constructs were generated by *Agrobacterium*-mediated transformation of leaf explants as described previously¹⁷. Different graft combinations were obtained as previously described¹⁷ and their tuberization phenotype was scored under long-day conditions. Tuber formation in *StSP6A* RNAi and *StSP3D* RNAi lines was analysed by growing the plants under long-day greenhouse conditions until a ten-leaf stage, before transferring them to short-day conditions. For flowering studies, plants were grown under high-irradiance ($>200 \mu\text{E s}^{-1} \text{m}^{-2}$) long-day conditions. At least nine replicates of each line were used for these studies.

Arabidopsis constans-1 (co-1) and *flowering locus t-1 (ft-1)* mutants in the Columbia (Col-0) and Landsberg erecta (Ler) backgrounds were transformed using the floral dip method. Flowering time was measured as the number of rosette leaves at floral initiation under long-day conditions, in at least ten individuals.

Plasmid constructs. The *rolC::Hd3a-GFP* construct has been described elsewhere¹⁸. Transcripts corresponding to the *StSP6A*, *StSP3D* and *StCO* genes were amplified using primers designed on the tomato sequences (AY186737, AY186735 and AY490253, respectively). The *StSP6A* overexpression construct was generated by amplifying the protein-coding region and then inserting the PCR product into the *Sma*I site of the pBinAR binary vector, between the 35S promoter and the *ocs* terminator. To silence the *StSP6A*, *StSP3D* and *StCO* transcripts, Andigena plants were transformed with RNAi constructs designed on the non-conserved region of the genes. The amplification products were cloned into the pENTR/D-TOPO plasmid (Invitrogen) and inserted in opposite orientations by recombination with the LR Clonase II enzyme (Gateway Technology, Invitrogen) into the pBIN19RNAi destination vector. The pBIN19RNAi interference vector was generated by partial *Xba*I/*Hind*III digestion of the pH7GWGW2(II) plasmid (<http://gateway.psb.ugent.be/>) and insertion of the Gateway cassette fragment into the same restriction sites of the pBIN19 binary vector. The *StCO* overexpression construct was generated by amplifying and cloning the protein-coding region into the pENTR/D-TOPO plasmid (Invitrogen), and further insertion by recombination into the pGWB2 destination vector³¹. To generate the *StSP6A* ethanol-inducible lines, the pBinSRNA-GW destination vector was created by insertion of the blunt-ended pAlcA-R1-R2-t35S cassette from the AlcAP-GW pGreen vector (gift of Patrick Laufs, INRA), generated by *Xba*I digestion, into the blunted *Hind*III sites of the binary vector pBinSRNACatN. The *StSP6A* coding region was then introduced in this destination vector by LR Clonase recombination. A list of the primer sets used to generate these constructs is shown in Supplementary Table 3.

Real-time and semiquantitative RT-PCR analyses. Expression of potato *FT*- and *TFL1*-like family members was quantified by real-time PCR. First-strand complementary DNA was synthesized from 2 μg total RNA and 1 μl of the reaction used for real-time gene expression analysis with the SYBR Green PCR master mix (Applied Biosystems). Quantitative PCR was performed using the Power SYBR Green PCR Master Mix (Applied Biosystems) on an ABI7500 Real-Time PCR System (Applied Biosystems), following the manufacturer's instructions. Primer pairs were specifically designed for each gene using PRIMER EXPRESS 3.0 software (Applied Biosystems) and probed for high-efficiency amplification under standard quantitative PCR conditions. All reactions were carried out at least in two independent biological replicates. The Pfaffl³² method was used for relative quantification of gene expression. Direct $2^{-\Delta C_T}$, where $\Delta C_T = C_T$ (target gene) $- C_T$ (actin) (ref. 33), was used to generate the tissue-specific heat map in Fig. 1e. The comparative critical threshold ($2^{-\Delta\Delta C_T}$) method³⁴ was used to analyse relative *Hd3a* and *StSP6A* expression levels (Supplementary Figs 1 and 10).

Identical procedures were used for semiquantitative RT-PCR, except that amplification was conducted in a Peltier thermal cycler (PTC-200, MJ Research) and one-tenth of the complementary DNA reaction was used for actin amplification. For sets of specific primers and product lengths, see Supplementary Table 3.

Microarray sample hybridization and analysis. Stolons of wild-type plants were collected in long days and two, six and eight days after transferring the plants to

short days. Time-course profiling analyses of short-day-induced stolons were performed as described previously³⁵.

Arrays of *StSP6A* overexpression versus wild type (sampled in long days) and *StSP6A* RNAi stolons versus wild type (sampled in short days, day 6) were performed with samples from three independent lines. Samples were hybridized against POCI (Potato Oligo Chip Initiative) microarrays, and background correction and normalization were performed using LIMMA^{36,37}. The obtained data was statistically checked (false discovery rate, <0.05) and genes with a log ratio change of $+2$ or -2 for *StSP6A* RNAi and $+1.8$ or -1.8 for *StSP6Aox* plants were selected. The hierarchical cluster of genes present in both experiments was calculated, compared with the short-day profile and represented using the TIGR MEV free software (<http://www.tm4.org/mev/>). Genes found to be differentially expressed are listed in Supplementary Table 1.

Immunoblot analysis. Total protein extracts from stolon tissues were obtained in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Na deoxycholate, 0.5% Triton X-100, 1 mM PMSF and protease inhibitors). For analysis of protein graft transmissibility, total extracts were incubated overnight with 10 μl of an anti-GFP affinity matrix (MBL). After extensive washing, unbound and bound proteins were separated by 10% SDS-polyacrylamide gel electrophoresis, blotted onto nitrocellulose and probed with an anti-GFP antibody (Roche). Immunoreactive proteins were detected with the SuperSignal West Pico Chemiluminescence kit (Pierce).

Microscopy. GFP fluorescence was observed on longitudinal stolon sections (150 μm) obtained with a vibratome (PELCO 101). Fluorescence was excited with a 488-nm argon laser and emission images were collected in the 500–600-nm range using a Leica TCS SP5 spectral confocal microscope.

For scanning electron microscopy, potato stolons and apical meristems were frozen in an Oxford CT 1500 cryosystem (Oxford Instruments), sublimated under vacuum and observed using a DIOL JSM 5410 electronic microscope operating at 10 kV.

Phylogenetic analysis. Full-length protein sequences, except sequences of the potato *FT*- and *TFL1*-like family members, were obtained from GenBank. Sequences of potato *FT*- and *TFL1*-like family members were obtained from the Potato Genome Sequencing Consortium Data Release²⁰ (<http://potatogenomics.plantbiology.msu.edu/>) by TBLASTX using as a query previously described tomato sequences¹⁹ (*SlSP*, no. U84140; *SlSP6A*, no. AY186737; *SlSP3D*, no. AY186735; *SlSP9D*, no. AY186738; *SlSP2I*, no. AY186734; and *SlSP5G*, no. AY186736). The best genome matches (Supplementary Table 2) were downloaded and open reading frames were predicted using FGENESH (<http://linux1.softberry.com/berry.phtml>) and ORF Finder (<http://www.ncbi.nlm.nih.gov/projects/gorf/>). A counterpart to *SlSP2I* was not found in the potato genome. Phylogenetic analyses were conducted using MEGA4³⁸. Sequences were aligned with the COBALT program (<http://www.ncbi.nlm.nih.gov/tools/cobalt/>) and their evolutionary relationship inferred using the neighbour-joining method³⁹. All ambiguous positions were removed for each sequence pair. The evolutionary distances were computed using the Poisson correction method and the bootstrap consensus tree was inferred from 1,000 replicates. The accession numbers for the corresponding genes are indicated in the tree (Supplementary Fig. 2).

- Nakagawa, T. *et al.* Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* **104**, 34–41 (2007).
- Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
- Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C_T method. *Nature Protocols* **3**, 1101–1108 (2008).
- Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**, 402–408 (2001).
- Kloosterman, B. *et al.* Genes driving potato tuber initiation and growth: identification based on transcriptional changes using the POCI array. *Funct. Integr. Genomics* **8**, 329–340 (2008).
- Smyth, G. K. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, 3 (2004).
- Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31**, 265–273 (2003).
- Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599 (2007).
- Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).

Peripheral SMN restoration is essential for long-term rescue of a severe spinal muscular atrophy mouse model

Yimin Hua¹, Kentaro Sahashi¹, Frank Rigo², Gene Hung², Guy Horev¹, C. Frank Bennett² & Adrian R. Krainer¹

Spinal muscular atrophy (SMA) is a motor neuron disease and the leading genetic cause of infant mortality; it results from loss-of-function mutations in the survival motor neuron 1 (*SMN1*) gene¹. Humans have a paralogue, *SMN2*, whose exon 7 is predominantly skipped², but the limited amount of functional, full-length SMN protein expressed from *SMN2* cannot fully compensate for a lack of *SMN1*. SMN is important for the biogenesis of spliceosomal small nuclear ribonucleoprotein particles³, but downstream splicing targets involved in pathogenesis remain elusive. There is no effective SMA treatment, but SMN restoration in spinal cord motor neurons is thought to be necessary and sufficient⁴. Non-central nervous system (CNS) pathologies, including cardiovascular defects, were recently reported in severe SMA mouse models and patients^{5–8}, reflecting autonomic dysfunction or direct effects in cardiac tissues. Here we compared systemic versus CNS restoration of SMN in a severe mouse model^{9,10}. We used an antisense oligonucleotide (ASO), ASO-10-27, that effectively corrects *SMN2* splicing and restores SMN expression in motor neurons after intracerebroventricular injection^{11,12}. Systemic administration of ASO-10-27 to neonates robustly rescued severe SMA mice, much more effectively than intracerebroventricular administration; subcutaneous injections extended the median lifespan by 25 fold. Furthermore, neonatal SMA mice had decreased hepatic *Igfals* expression, leading to a pronounced reduction in circulating insulin-like growth factor 1 (IGF1), and ASO-10-27 treatment restored IGF1 to normal levels. These results suggest that the liver is important in SMA pathogenesis, underscoring the importance of SMN in peripheral tissues, and demonstrate the efficacy of a promising drug candidate.

To compare the effectiveness of ASO-10-27 delivered centrally versus systemically, we administered an intracerebroventricular (ICV) injection of 20 μg ASO-10-27 on postnatal day 1 (P1) to increase SMN expression in CNS tissues, or we administered a subcutaneous (SC) injection of the ASO on two separate days at 50 μg per g of body weight ($\mu\text{g g}^{-1}$), between P0 and P3 (two doses). These doses were based on our previous studies with this ASO^{11,13}. We also evaluated combined ICV and SC injections, as well as repeated SC injections (Supplementary Table 1). Control heterozygous mice (*Smn*^{+/-} *SMN2*^{+/-}) that received ICV and/or SC ASO-10-27 injections had normal survival and behaviour. Severe SMA mice (*Smn*^{-/-} *SMN2*^{+/-}) that received ICV and/or SC saline injections survived for 1–2 weeks, with a median survival time of ~10 days, similar to untreated mice (Fig. 1a, Supplementary Figs 1a and 2a, and Supplementary Movie 1). Delivery of the ASO only into the CNS efficiently corrected *SMN2* exon 7 splicing in the spinal cord and led to a striking increase in SMN protein levels, but modestly extended the median survival to 16 days, with a single pup surviving for 1 month (Fig. 1a–c and Supplementary Fig. 2b–d). In marked contrast, systemic treatment with two SC injections resulted in a median survival of 108

days (Fig. 1d). Combining ICV and SC injections of the ASO further increased the median survival to 173 days, and two additional SC injections on P5 and P7, after the initial SC injections at P0–P3, extended the median survival to 137 days (Fig. 1d).

Treated SMA mice varied in size from runts to comparable to their heterozygous littermates; their average weight was low, and their tails were much shorter than normal (Supplementary Figs 3 and 4). The surviving runts slowly gained weight, reaching ~18 g at ~3 months.

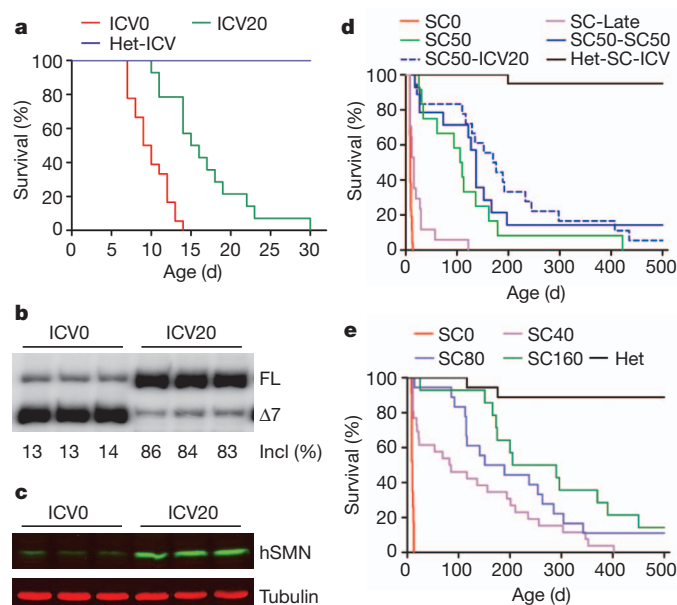


Figure 1 | Systemic versus ICV ASO-10-27 injections in SMA mice.

a, Survival curves for mice after ICV administration of ASO-10-27 on P1. Administration of 20 μg ASO-10-27 (ICV20, $n = 14$) or saline (ICV0, $n = 18$) resulted in mean survival times of 17 and 10 days (d), respectively ($P < 0.001$). ASO-10-27-treated heterozygotes (Het-ICV, $n = 15$) served as controls. **b**, **c**, Spinal cord RNA and protein samples ($n = 3$) were analysed on P7 by using radioactive RT-PCR (**b**) or immunoblotting with a monoclonal antibody specific for human SMN (hSMN) (**c**). $\Delta 7$, exon 7-skipped mRNA; FL, full-length mRNA; incl, exon 7 inclusion; % incl = $100 \times \Delta 7 / (\text{FL} + \Delta 7)$. **d**, Survival curves after SC administration of saline (SC0, $n = 26$) or ASO-10-27 (SC50, $n = 12$) twice between P0 and P3. SC50-SC50 ($n = 14$) mice received two additional SC injections on P5 and P7. Het-SC-ICV ($n = 13$) and SC50-ICV20 ($n = 18$) were heterozygous and SMA mice, respectively, that received combined P1 ICV and P0–P3 SC injections. SC-Late ($n = 17$) were SMA mice that received only two SC injections, on P5 and P7. Each SC injection dose was 50 $\mu\text{g g}^{-1}$ body weight. $P < 0.0001$ for all groups versus SC0 except for SC-Late, $P < 0.05$. **e**, Dose-dependent survival after two SC injections at P0–P3 with 40 (SC40, $n = 26$), 80 (SC80, $n = 18$) or 160 (SC160, $n = 14$) $\mu\text{g g}^{-1}$ of ASO-10-27. Saline-treated SMA (SC0, $n = 23$) or heterozygous mice (Het, $n = 18$) served as controls. $P < 0.0001$ for all groups versus SC0.

¹Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA. ²Isis Pharmaceuticals, 2855 Gazelle Court, Carlsbad, California 92010, USA.

Most rescued SMA mice could run and climb normally; however, their tails and ears developed necrosis and were gradually lost, resembling the phenotype of type III SMA mice (Supplementary Fig. 3e, f). Additional delivery of the ASO either by ICV injection on P1 or repeat SC injections on P5 and P7 delayed necrosis (Supplementary Fig. 3g).

To further characterize the effects of the ASO administered systemically, we carried out a dose-response study with 0 (SC0), 40 (SC40), 80 (SC80) and 160 (SC160) $\mu\text{g g}^{-1}$ ASO-10-27 as an SC injection, given twice between P0 and P3. Systemic treatment with the ASO resulted in a dose-dependent increase in survival (Fig. 1e), with the median survival increasing from 10 days to 84, 170 and 248 days, respectively. At the highest dose tested, the ASO given systemically resulted in long-term survival comparable to the best results achieved by adeno-associated virus expression of the SMN protein in a slightly less severe mouse model^{14–16}. Remarkably, 2 of 14 mice in the SC160 group and 2 of 18 in the SC80 group are still alive and active after >500 days. A similar survival benefit was achieved by intraperitoneal administration (Supplementary Fig. 5). There was no significant difference in weight gain among the three SC-dosing groups; however, mice in the SC160 group had significantly longer tails (Supplementary Fig. 6a–e). We also observed dose-dependent rescue of ear and tail necrosis and dose-dependent delays in the development of cataracts and rectal prolapse (Supplementary Fig. 6f–h). Administration of the ASO in two doses on P5 and P7 resulted in a modest increase in survival, compared with earlier treatment (between P0 and P3), emphasizing the importance of early postnatal therapeutic intervention (Fig. 1d).

To examine *SMN2* splicing changes in various tissues after SC injection of the ASO, we performed reverse transcription followed by PCR (RT-PCR) on RNA samples from P7 mice. We detected a dose-dependent increase in exon 7 inclusion in the spinal cord, brain, liver, heart, kidneys and skeletal muscle, with the strongest effect occurring in the liver and the weakest in the kidneys (Fig. 2a, b and Supplementary Fig. 7a, b). By contrast, ICV administration of the ASO resulted in a much more robust change in exon 7 inclusion in the brain and spinal cord tissues but had very limited effects in peripheral tissues (Supplementary Fig. 8a). Immunoblotting of spinal cord, liver and heart tissue samples from mice treated by SC administration showed a corresponding increase in full-length SMN protein (Fig. 2c and Supplementary Figs 7c and 8b). Exon 7 inclusion in the liver significantly decreased after P30 (Fig. 2d, e and Supplementary Fig. 9), consistent with the measured half-life of the ASO being 22 days in the liver (data not shown). These data suggest that transiently increasing SMN expression in peripheral tissues during the first few weeks of life has a profound effect on long-term survival of severe SMA mice.

The *SMN2*-splicing changes were consistent with the ASO distribution, as assayed by immunohistochemistry, with the apparent exception of the kidneys; however, in the kidneys, most of the ASO had not been internalized by the cells (Supplementary Figs 10 and 11). We also observed some of the ASO accumulating in spinal cord motor neurons (Supplementary Figs 10 and 11). The limited distribution of the ASO and the moderate *SMN2*-splicing changes in the CNS after systemic administration probably reflect incomplete closure of the blood–brain barrier in neonates¹⁷ and/or retrograde transport of the ASO. However, we detected strong cytoplasmic SMN staining and/or a pronounced increase in gem number in spinal cord motor neurons after ICV injection of 20 μg ASO-10-27 but not after two SC injections of ASO-10-27 at 80 $\mu\text{g g}^{-1}$ between P0 and P3, a dosage that substantially rescued the severe SMA mice (Supplementary Figs 12 and 13). Therefore, the effect on splicing in the CNS after systemic administration probably contributes to the extended survival, which is consistent with the combined ICV and SC treatment resulting in even better survival than SC administration alone (Fig. 1d). However, the striking effects of systemic administration on survival in this severe mouse model cannot be explained solely by a direct effect on *SMN2* splicing in the CNS.

The rescue of severe SMA mice by systemic administration of, for example, histone deacetylase inhibitors or adeno-associated virus

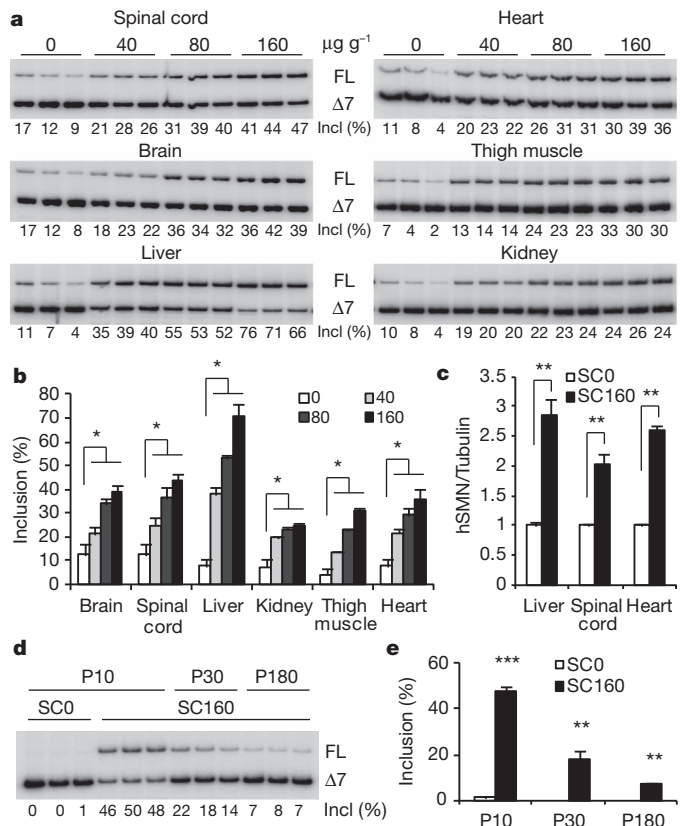


Figure 2 | *SMN2* splicing and protein expression in mouse tissues after SC injection of ASO-10-27. **a**, Tissue RNA samples from SMA mice were analysed by radioactive RT-PCR at P7 after two SC injections between P0 and P3 with 0, 40, 80 or 160 $\mu\text{g g}^{-1}$ body weight ASO-10-27. **b**, Histogram of exon 7 inclusion data from panel **a** ($n = 3$). **c**, Protein samples from P7 SMA mice ($n = 3$) that had been treated with 160 $\mu\text{g g}^{-1}$ body weight ASO-10-27 were analysed by immunoblotting with a monoclonal antibody specific for human SMN (see also Supplementary Fig. 8b). **d**, RT-PCR of liver RNA from P10, P30 and P180 SMA mice, showing the decreasing effect of ASO-10-27 over time. **e**, Histogram of data from panel **d**. **b, c, e**, Data are presented as mean \pm s.d. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.0001$; all compared with saline controls.

vectors, has been attributed to the ability of these agents to cross the blood–brain barrier^{10,15}. However, our data indicate that SMN restoration in peripheral tissues, in combination with partial restoration in the CNS, can achieve efficient rescue of severe SMA mice.

In mice that had been treated systemically with 160 $\mu\text{g g}^{-1}$ ASO-10-27 and sacrificed on P9, histological examination of tissues or organs associated with SMA revealed striking improvements, consistent with the markedly increased survival of mice in the SC160 group. The α -motor neuron counts in the spinal cords of these mice were comparable to those of the control heterozygous littermates, and the mean area of muscle fibre cross-sections was >80% of that of heterozygotes (Fig. 3a, b and Supplementary Fig. 14a). Likewise, the heart weight and the thickness of the interventricular septum and the left ventricular wall were similar in mice in the SC160 group and their heterozygous littermates (Fig. 3c, d and Supplementary Fig. 14b). Finally, staining of the neuromuscular junctions (NMJs) showed that NMJ integrity was similar in mice in the SC160 group and their heterozygous littermates (Fig. 3e).

Most of the mice that were treated systemically with the ASO showed no overt signs of motor dysfunction (Supplementary Movie 2 and Supplementary Table 2). We used three tests to evaluate behaviour and motor function. The first was a rotarod test, which requires limb muscle strength, as well as balance and coordination. Three-month-old mice in the SC80 and SC160 groups could stay on the rotating rod for ~ 12 s: that is, for less time than the control heterozygotes but for longer

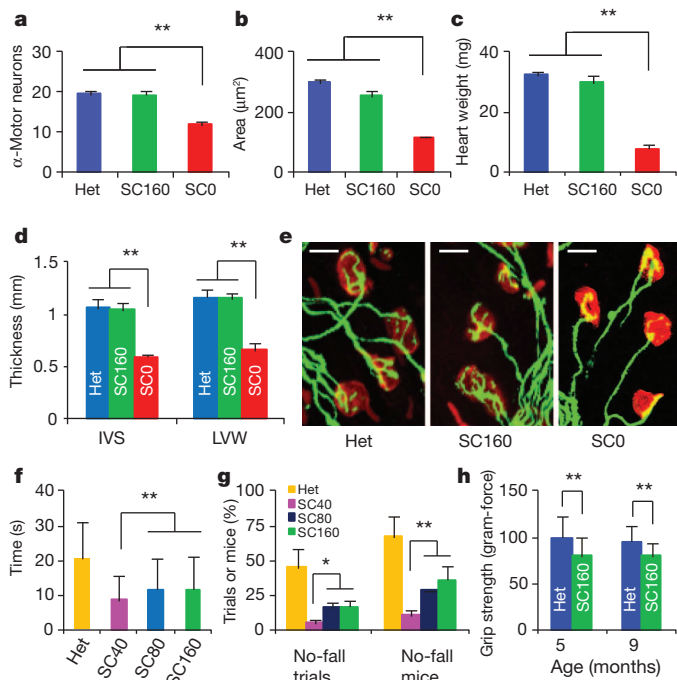


Figure 3 | Evaluation of affected tissues and motor function. Tissues from three groups of P9 mice were stained with haematoxylin and eosin: SMA mice that had been treated with ASO-10-27 (SC160, two SC injections at $160 \mu\text{g g}^{-1}$ body weight at P0–P3, $n = 6$), saline controls (SC0, $n = 6$) and untreated heterozygotes (Het, $n = 6$) (see also Supplementary Fig. 14). Saline-treated mice were ambulant at P9 and were expected to live for another 3–5 days. α -Motor neuron counts in each cross-section of the L1–L2 spinal cord (a), mean fibre cross-sectional area (for a total of 200 fibres) of the rectus femoris muscle (b), heart weight (c) and thickness of the heart interventricular septum (IVS) and left ventricular wall (LVW) (d) significantly improved in ASO-10-27-treated mice. e, The arborization complexity of NMJs was restored in ASO-10-27-treated mice (red, endplates; green, neurofilament medium). Scale bar, $10 \mu\text{m}$. f, g, P90 SC40 ($n = 12$, 124 trials), SC80 ($n = 13$, 137 trials), SC160 ($n = 11$, 117 trials) and untreated heterozygous (Het, $n = 12$, 135 trials) mice were tested three to five times per day for 3 days on a rotarod, using an acceleration profile. The mean times for staying on the spinning rod (f) and the percentage of no-fall trials and of mice with ≥ 1 no-fall trial (g) are shown. h, The grip strength (gram-force) of SC160 mice ($n = 6$) evaluated at 5 and 9 months reached $\sim 80\%$ of that of heterozygous mice ($n = 6$). a–d, f–h, Data are presented as mean + s.d. *, $P < 0.05$; **, $P < 0.01$.

than mice in the SC40 group. Some ASO-10-27-treated mice passed a 30 s acceleration-profile test that many of the heterozygotes failed (Fig. 3f, g and Supplementary Movie 3). Considering that SMA is a neuromuscular disease, this performance represents a remarkable phenotypic improvement. The second test evaluated muscle strength in mice from the SC160 group at 5 and 9 months. At both ages, the forelimb grip strength of treated SMA mice was $\sim 80\%$ that of the control heterozygous mice (Fig. 3h). The final test used HomeCageScan, a video-based platform for automated high-resolution behaviour analysis¹⁸. ASO-10-27-treated SMA mice performed various behaviours similarly to control heterozygous mice, except for rearing, suggesting some hindlimb weakness (Supplementary Fig. 15).

Two observations prompted us to examine the growth hormone (GH)–IGF1 axis. First, all severe SMA mice are small^{9,10,19}, reflecting growth retardation (Supplementary Figs 2a and 3c). Second, the major effect of SC injection of the ASO on *SMN2* splicing is in the liver, which contributes $\sim 75\%$ of the circulating IGF1 (ref. 20). Moreover, restoring IGF1 expression in the liver is sufficient to support normal postnatal growth of *Igf1*-null mice²⁰. IGF1 is a potent neurotrophic factor²¹ and is also involved in cardiac development and function²². An enzyme-linked immunosorbent assay (ELISA) of serum samples from SMA mice at P6–P9 showed that IGF1 was undetectable or present at greatly

reduced levels compared with the heterozygous controls; SC administration of the ASO restored IGF1 to normal levels in SMA mice (Fig. 4a).

RT-PCR showed that the level of hepatic *Igf1* messenger RNA was not reduced in SMA mice compared with the heterozygous controls and that it increased from P1 to P5 in both SMA and control mice (Fig. 4b). IGF-binding protein, acid labile subunit (IGFALS), which is postnatally stimulated by GH, binds to IGF1 and IGF-binding protein 3 (IGFBP3) to form a stable ternary complex, extending the half-life of IGF1 from 10 min to $>12 \text{ h}$ ²³. The inactivation of *Igfals* results in low levels of circulating IGF1 and IGFBP3, as well as impaired postnatal growth²³. RT-PCR revealed a marked reduction in the amount of *Igfals* mRNA in the liver of both P1 and P5 SMA mice compared with the heterozygous controls; moreover, administration of the ASO rescued *Igfals* expression (Fig. 4c, d). We conclude that the striking reduction in serum IGF1 levels in SMA mice is likely to be caused by decreased *Igfals* expression, which correlates with SMN deficiency and SMA progression.

Because *Igfals* expression is decreased on P1, when the pups are still healthy, we propose that the early deficiency in circulating IGF1 may be one of the factors that contribute to the pathogenesis of severe SMA mice (Supplementary Fig. 1b). Although in several mouse mutants, an impaired GH–IGF1 axis results in an increased lifespan²⁴, a severe lack of IGF1 may contribute to SMA progression, together with other defective factors. Consistent with our hypothesis, two recent studies have shown that a local increase in IGF1 in either the spinal cord or muscle increases the survival of severe SMA mice^{25,26}. Indeed, disruption of the IGF1 system is a common feature of neurodegenerative diseases, including Alzheimer's disease and amyotrophic lateral sclerosis (ALS)²⁷. *Igf1*-null mice also show some phenotypic similarity to SMA mice, such as small size and severe and generalized muscle dystrophy (including of the diaphragm and heart), with most of them dying at birth²⁸. Moreover, dysregulation of the IGF1 receptor and its downstream signalling pathway has been observed in patients with type I SMA²⁹. However, the results of IGF1 therapy for ALS are not consistent between mice and humans^{21,30}. In light of this inconsistency,

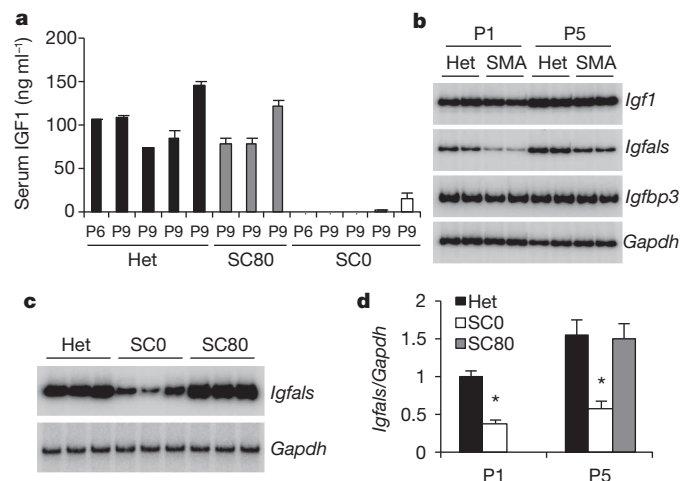


Figure 4 | The IGF1 system is disrupted in SMA mice. Treated SMA mice (SC80) received two SC injections of ASO-10-27 at $80 \mu\text{g g}^{-1}$ body weight between P0 and P3. a, IGF1 serum levels in P6 and P9 SMA mice (SC0), as measured by ELISA (mean of three measurements per sample), were strikingly lower than for their heterozygous littermates (Het) or treated SMA mice ($P < 0.001$ for all samples). b, Total liver RNA from P1 and P5 SMA mice and their heterozygous littermates was analysed by radioactive RT-PCR to measure *Igf1*, *Igfals* and *Igfbp3* expression, with *Gapdh* as a control. c, ASO-10-27 treatment restored hepatic *Igfals* expression at P5, as shown by radioactive RT-PCR. d, Quantification of hepatic *Igfals* expression ($n = 5$). *, $P < 0.01$ versus samples from heterozygous littermates or ASO-10-27-treated SMA mice. a, d, Data are presented as mean + s.d.

it will be crucial to determine the extent to which SMA mouse models accurately mimic human SMA. This will further refine our understanding of the mouse models and influence the development of therapeutics and clinical treatments for SMA.

METHODS SUMMARY

ASO-10-27 was synthesized as described previously¹¹ and dissolved in 0.9% saline. The severe SMA mouse model was generated from a type III SMA mouse model, as described previously^{9,10}. All mouse protocols were in accordance with the Cold Spring Harbor Laboratory's Institutional Animal Care and Use Committee guidelines. Treated mice (control and ASO-10-27) were provided with additional gel food. The procedures for neonatal ICV injection, tissue sample collection, RT-PCR, western blotting and the human-specific anti-SMN antibody (SMN-KH) were as described previously¹¹. The primers for the gene expression analysis are shown in Supplementary Table 3. Serum IGF1 levels were analysed with a Mouse/Rat IGF-I Quantikine ELISA kit (R&D Systems).

Mouse spinal cords, quadriceps and hearts were fixed and stained with haematoxylin and eosin as described previously¹². α -Motor neurons were counted in serial 10–20- μ m cross-sections of the lumbar (L1–L2) spinal cord. The muscle fibre cross-sectional area was calculated by using AxioVision LE velocity software. NMJ staining *in toto* was performed as described previously¹².

For the rotarod (AccuScan Instruments) test, a four-phase profile was used: phase 1, from 1 to 10 r.p.m. in 7.5 s; phase 2, from 10 to 0 r.p.m. in 7.5 s; phase 3, from 0 to 10 r.p.m. in 7.5 s in the opposite direction; and phase 4, from 10 to 0 r.p.m. in 7.5 s. A grip-strength meter (Columbus Instruments) was used for the gripping test. Mice were allowed to grasp a triangular bar with their forelimbs and were pulled back horizontally. The test was repeated five times for each mouse, and the highest value was recorded as the grip force for that animal.

Statistical significance was analysed by two-tailed Student's *t*-tests. Kaplan–Meier survival data were analysed with Mantel–Cox tests using the program GraphPad Prism.

Received 15 April; accepted 18 August 2011.

- Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80**, 155–165 (1995).
- Lorson, C. L., Rindt, H. & Shababi, M. Spinal muscular atrophy: mechanisms and therapeutic strategies. *Hum. Mol. Genet.* **19**, R111–R118 (2010).
- Burghes, A. H. & Beattie, C. E. Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nature Rev. Neurosci.* **10**, 597–609 (2009).
- Gavrilina, T. O. *et al.* Neuronal SMN expression corrects spinal muscular atrophy in severe SMA mice while muscle-specific SMN expression has no phenotypic effect. *Hum. Mol. Genet.* **17**, 1063–1075 (2008).
- Rudnik-Schoneborn, S. *et al.* Congenital heart disease is a feature of severe infantile spinal muscular atrophy. *J. Med. Genet.* **45**, 635–638 (2008).
- Bevan, A. K. *et al.* Early heart failure in the SMN Δ 7 model of spinal muscular atrophy and correction by postnatal scAAV9-SMN delivery. *Hum. Mol. Genet.* **19**, 3895–3905 (2010).
- Heier, C. R., Satta, R., Lutz, C. & DiDonato, C. J. Arrhythmia and cardiac defects are a feature of spinal muscular atrophy model mice. *Hum. Mol. Genet.* **19**, 3906–3918 (2010).
- Shababi, M. *et al.* Cardiac defects contribute to the pathology of spinal muscular atrophy models. *Hum. Mol. Genet.* **19**, 4059–4071 (2010).
- Gogliotti, R. G., Hammond, S. M., Lutz, C. & DiDonato, C. J. Molecular and phenotypic reassessment of an infrequently used mouse model for spinal muscular atrophy. *Biochem. Biophys. Res. Commun.* **391**, 517–522 (2010).
- Riessland, M. *et al.* SAHA ameliorates the SMA phenotype in two mouse models for spinal muscular atrophy. *Hum. Mol. Genet.* **19**, 1492–1506 (2010).
- Hua, Y. *et al.* Antisense correction of SMN2 splicing in the CNS rescues necrosis in a type III SMA mouse model. *Genes Dev.* **24**, 1634–1644 (2010).
- Passini, M. A. *et al.* Antisense oligonucleotides delivered to the mouse CNS ameliorate symptoms of severe spinal muscular atrophy. *Sci. Transl. Med.* **3**, 72ra18 (2011).
- Hua, Y., Vickers, T. A., Okunola, H. L., Bennett, C. F. & Krainer, A. R. Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. *Am. J. Hum. Genet.* **82**, 834–848 (2008).
- Passini, M. A. *et al.* CNS-targeted gene therapy improves survival and motor function in a mouse model of spinal muscular atrophy. *J. Clin. Invest.* **120**, 1253–1264 (2010).
- Foust, K. D. *et al.* Rescue of the spinal muscular atrophy phenotype in a mouse model by early postnatal delivery of SMN. *Nature Biotechnol.* **28**, 271–274 (2010).
- Dominguez, E. *et al.* Intravenous scAAV9 delivery of a codon-optimized SMN1 sequence rescues SMA mice. *Hum. Mol. Genet.* **20**, 681–693 (2011).
- Ek, C. J., Habgood, M. D., Dziegielewska, K. M. & Saunders, N. R. Structural characteristics and barrier properties of the choroid plexuses in developing brain of the opossum (*Monodelphis domestica*). *J. Comp. Neurol.* **460**, 451–464 (2003).
- Steele, A. D., Jackson, W. S., King, O. D. & Lindquist, S. The power of automated high-resolution behavior analysis revealed by its application to mouse models of Huntington's and prion diseases. *Proc. Natl Acad. Sci. USA* **104**, 1983–1988 (2007).
- Park, G. H., Kariya, S. & Monani, U. R. Spinal muscular atrophy: new and emerging insights from model mice. *Curr. Neurol. Neurosci. Rep.* **10**, 108–117 (2010).
- Wu, Y., Sun, H., Yakar, S. & LeRoith, D. Elevated levels of insulin-like growth factor (IGF)-I in serum rescue the severe growth retardation of IGF-I null mice. *Endocrinology* **150**, 4395–4403 (2009).
- Kaspar, B. K., Llado, J., Sherkat, N., Rothstein, J. D. & Gage, F. H. Retrograde viral delivery of IGF-1 prolongs survival in a mouse ALS model. *Science* **301**, 839–842 (2003).
- Colao, A. The GH-IGF-I axis and the cardiovascular system: clinical implications. *Clin. Endocrinol.* **69**, 347–358 (2008).
- Domené, H. M. *et al.* Human acid-labile subunit deficiency: clinical, endocrine and metabolic consequences. *Horm. Res.* **72**, 129–141 (2009).
- Kenyon, C. The plasticity of aging: insights from long-lived mutants. *Cell* **120**, 449–460 (2005).
- Shababi, M., Glascock, J. & Lorson, C. L. Combination of SMN trans-splicing and a neurotrophic factor increases the life span and body mass in a severe model of spinal muscular atrophy. *Hum. Gene Ther.* **22**, 135–144 (2011).
- Bosch-Marce, M. *et al.* Increased IGF-1 in muscle modulates the phenotype of severe SMA mice. *Hum. Mol. Genet.* **20**, 1844–1853 (2011).
- Trejo, J. L., Carro, E., Garcia-Galloway, E. & Torres-Aleman, I. Role of insulin-like growth factor I signaling in neurodegenerative diseases. *J. Mol. Med.* **82**, 156–162 (2004).
- Powell-Braxton, L. *et al.* IGF-I is required for normal embryonic growth in mice. *Genes Dev.* **7**, 2609–2617 (1993).
- Millino, C. *et al.* Different atrophy-hypertrophy transcription pathways in muscles affected by severe and mild spinal muscular atrophy. *BMC Med.* **7**, 14 (2009).
- Sorenson, E. J. *et al.* Subcutaneous IGF-1 is not beneficial in 2-year ALS trial. *Neurology* **71**, 1770–1775 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We gratefully acknowledge support from the Muscular Dystrophy Association, the National Institute of General Medical Sciences and St. Giles Foundation. We thank J. Bu and M. Passini for protocols and advice on NMJ staining, and S. Hearn for assistance with microscope imaging.

Author Contributions Y.H., A.R.K. and C.F.B. designed the study and wrote the paper. Y.H., K.S., F.R., G. Hung and G. Horev carried out the experiments and analysed the data. All authors read the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.R.K. (Krainer@cshl.edu).

Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy

Mariko Taniguchi-Ikeda^{1,2*}, Kazuhiro Kobayashi^{1*}, Motoi Kanagawa¹, Chih-chieh Yu¹, Kouhei Mori¹, Tetsuya Oda¹, Atsushi Kuga¹, Hiroki Kurahashi³, Hasan O. Akman⁴, Salvatore DiMauro⁴, Ryuji Kaji⁵, Toshifumi Yokota⁶, Shin'ichi Takeda⁷ & Tatsushi Toda¹

Fukuyama muscular dystrophy (FCMD; MIM253800), one of the most common autosomal recessive disorders in Japan, was the first human disease found to result from ancestral insertion of a SINE-VNTR-*Alu* (SVA) retrotransposon into a causative gene^{1–3}. In FCMD, the SVA insertion occurs in the 3' untranslated region (UTR) of the *fukutin* gene. The pathogenic mechanism for FCMD is unknown, and no effective clinical treatments exist. Here we show that aberrant messenger RNA (mRNA) splicing, induced by SVA exon-trapping, underlies the molecular pathogenesis of FCMD. Quantitative mRNA analysis pinpointed a region that was missing from transcripts in patients with FCMD. This region spans part of the 3' end of the *fukutin* coding region, a proximal part of the 3' UTR and the SVA insertion. Correspondingly, *fukutin* mRNA transcripts in patients with FCMD and SVA knock-in model mice were shorter than the expected length. Sequence analysis revealed an abnormal splicing event, provoked by a strong acceptor site in SVA and a rare alternative donor site in *fukutin* exon 10. The resulting product truncates the *fukutin* carboxy (C) terminus and adds 129 amino acids encoded by the SVA. Introduction of antisense oligonucleotides (AONs) targeting the splice acceptor, the predicted exonic splicing enhancer and the intronic splicing enhancer prevented pathogenic exon-trapping by SVA in cells of patients with FCMD and model mice, rescuing normal *fukutin* mRNA expression and protein production. AON treatment also restored *fukutin* functions, including O-glycosylation of α -dystroglycan (α -DG) and laminin binding by α -DG. Moreover, we observe exon-trapping in other SVA insertions associated with disease (hypercholesterolemia⁴, neutral lipid storage disease⁵) and human-specific SVA insertion in a novel gene. Thus, although splicing into SVA is known^{6–8}, we have discovered in human disease a role for SVA-mediated exon-trapping and demonstrated the promise of splicing modulation therapy as the first radical clinical treatment for FCMD and other SVA-mediated diseases.

FCMD (incidence 1/34,000 births) shares phenotypic similarities with other severe muscular dystrophies, including muscle-eye-brain disease and Walker-Warburg syndrome. All show deficiencies in O-glycosylation of α -DG, an extracellular protein anchored on the plasma membrane. Insufficient O-glycosylation interferes with the ability of α -DG to interact with extracellular matrix proteins such as laminin^{9,10}. For this reason, FCMD, muscle-eye-brain disease and Walker-Warburg syndrome are categorized as ' α -dystroglycanopathies (α -DGopathy)¹⁰'; so far, no effective treatments exist for these conditions. SVA is a hominid-specific, composite non-coding retrotransposon that contains SINE (short interspersed sequence), VNTR (variable number of tandem repeat), and *Alu* sequences. It is still active

in humans, polymorphic and mobilized by the human LINE-1 *in trans*^{6,11–15}.

In previous work, we showed that *fukutin* mRNA (10 exons, 7.4- and 6.4-kilobase (kb) cDNAs in size with two poly-A sites, 461-amino-acid protein with calculated molecular mass of 53.7 kDa) was not detectable by northern blot analysis in patients with FCMD carrying the SVA insertion². To investigate the aetiology of this decreased expression, we have now analysed whole *fukutin* mRNA in lymphoblasts from patients with FCMD using quantitative PCR with reverse transcription (qRT-PCR). PCR products corresponding to the protein-coding region of *fukutin*, as well as those including sequences in the distal part of the 3' UTR (and thus downstream of the SVA insertion), were similar in abundance to those from an unaffected control (Fig. 1a). However, products located at sequence positions within the 3' UTR were markedly decreased relative to the control. From these results and along with previous reports of many 3' and 5' splice sites within SVA elements^{6–8}, we hypothesized that abnormal splicing occurs somewhere between the end of the *fukutin* protein-coding region and the SVA insertion.

We then performed long-range RT-PCR using primers that flank the region corresponding to decreased expression. In patients with FCMD, we detected a single 3-kb PCR product, which is shorter than the 5-kb product seen in the normal control (Fig. 1b). This observation was consistent in several tissue types from patients with FCMD (Supplementary Fig. 1). PCR from genomic DNA produced an 8-kb product in patients with FCMD, compared with a 5-kb product in the control (Fig. 1b). Sequence analysis of the 3-kb product from FCMD cDNA revealed a splicing event (Supplementary Fig. 2). This event generates a new donor-side breakpoint within the final coding exon (exon 10), located 116 base pairs (bp) upstream from the authentic stop codon. A rare alternative donor site at that position is activated and trapped by an alternative acceptor site located within the inserted SVA, creating an additional and aberrant exonic sequence (exon 11) (Fig. 1c). The acceptor-side breakpoint is located 274 bp downstream from the start of the SVA insertion, between ag and TC (Fig. 1c). The acceptor site has not been described in the previous reports of SVA splicing^{6,7}. This location is preceded by a pyrimidine-rich stretch, the SVA (TCTCCC)₄₁ hexamer at the 5' end of the SVA element, with a possible favourable branch point. Predicted exonic splicing enhancer sites occur around 70 bp downstream from the new acceptor site. We confirmed that the aberrant splicing event can be abolished by replacing AG with GG at the acceptor junction in cultured cells transfected with a *fukutin* construct carrying SVA insertion (Supplementary Fig. 3). *Fukutin* expression was not altered by cycloheximide treatment, indicating that the transcript was not subject to nonsense-mediated mRNA decay, possibly because this exon-trapping occurred within the last

¹Division of Neurology/Molecular Brain Science, Kobe University Graduate School of Medicine, Kobe 650-0017, Japan. ²Division of General Pediatrics, Kobe University Graduate School of Medicine, Kobe 650-0017, Japan. ³Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Aichi 470-1192, Japan. ⁴Department of Neurology, Columbia University Medical Center, New York, NY 10032, USA. ⁵Department of Clinical Neuroscience, The University of Tokushima Graduate School, Tokushima 770-8503, Japan. ⁶Department of Medical Genetics, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB T6G 2H7, Canada. ⁷Department of Molecular Therapy, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo 187-8502, Japan.

*These authors contributed equally to this work.

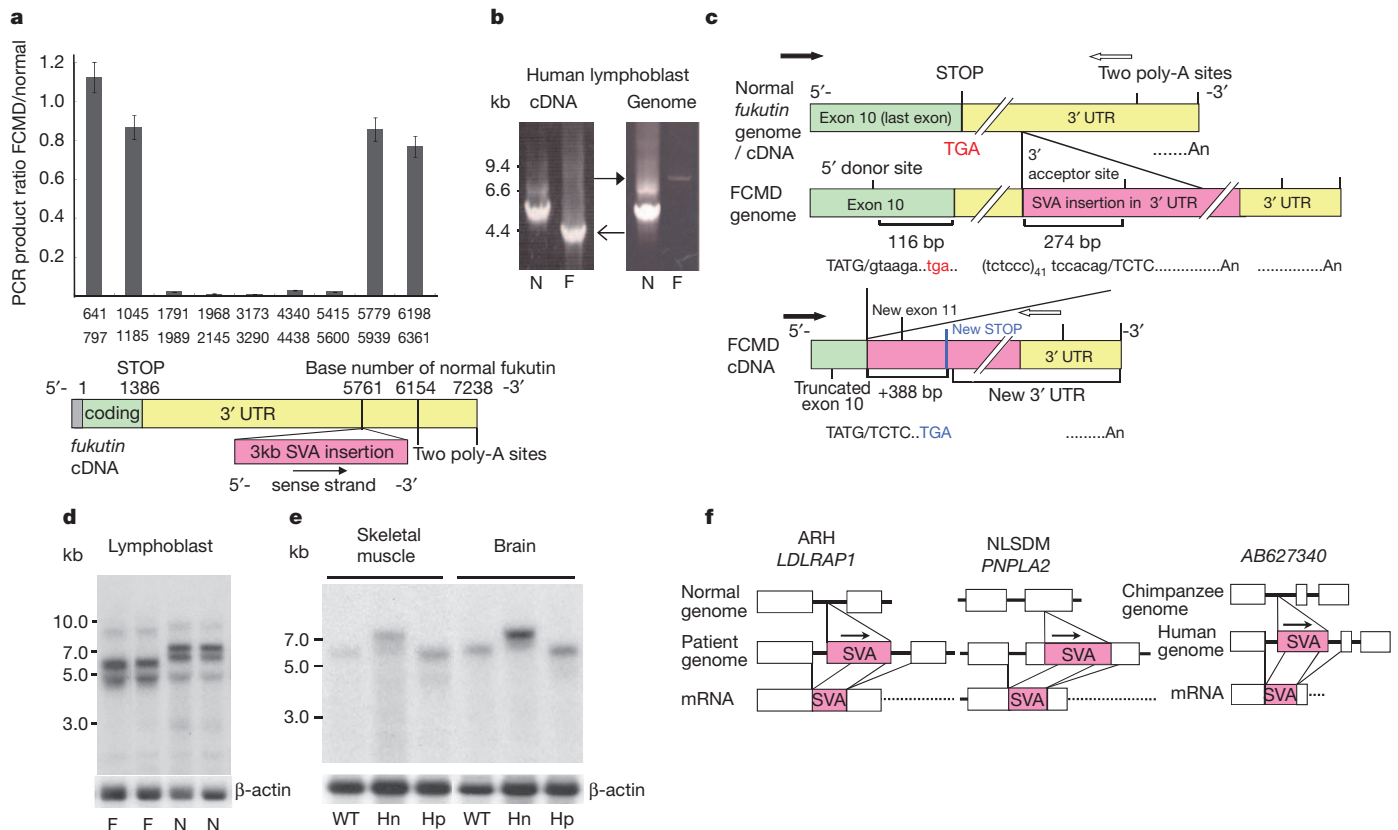


Figure 1 | An SVA retrotransposal insertion induces abnormal splicing in FCMD. **a**, Expression analysis of various regions of *fukutin* mRNA in lymphoblasts. Grey bar, the ratio of RT-PCR product in patients with FCMD relative to the normal control; numbers on the x axis, nucleotide positions of both forward and reverse primers in *fukutin*. Error bars, s.e.m. **b**, Long-range PCR using primers flanking the expression-decreasing area (nucleotide position 1,061–5,941) detected a 3-kb PCR product in FCMD lymphoblast cDNA (open arrow) and an 8-kb product in FCMD genomic DNA (filled arrow). In the normal control, cDNA and genomic DNA both showed 5-kb PCR products. The 8-kb band was weak, probably because VNTR region of

exon, and the new stop codon exists downstream of the new last exon–exon junction (Supplementary Fig. 4).

We have recently generated knock-in mice that carry a humanized *fukutin* exon 10, which either includes (Hp allele) or excludes (Hn allele) the SVA insertion, and bred these strains with heterozygous *fukutin* knockout mice to obtain compound heterozygotes (Hp/–)¹⁶. Knock-in mice that are homozygous (Hp/Hp) and compound heterozygous (Hp/–) are representative of the human FCMD alleles. These mice exhibit hypoglycosylation of α -DG in skeletal muscle, which is the most significant characteristic in α -DGpathy¹⁶. Quantitative RT-PCR in various tissues from Hp/Hp mice revealed an aberrant splicing pattern identical to that seen in human patients (Supplementary Fig. 5). Northern blot analysis detected abnormally spliced *fukutin* mRNA species at the expected sizes of 5.6 and 4.6 kb in patients with FCMD, whereas the normal *fukutin* mRNAs appeared at 7.4 and 6.4 kb (Fig. 1d and Methods). We replicated these results in the knock-in model mice (Fig. 1e and Supplementary Fig. 6a). The consistent observations between patients with FCMD and knock-in model mice lead us to conclude that a splicing abnormality underlies the pathogenesis of FCMD.

Abnormal splicing excises the authentic stop codon and produces another stop codon located 388 bp downstream from the 5' side of the new exon 11 (Fig. 1c). The predicted protein lacks the C-terminal 38 amino acids of *fukutin*, instead containing 129 amino acids derived from the SVA sequence (Supplementary Fig. 7). Endogenous *fukutin* is scarce and difficult to detect; however, we were able to identify both

SVA is GC-rich (82%). **c**, Representation of genomic DNA and cDNA in FCMD. Black and white arrows, forward and reverse sequencing primers. The intronic sequence in FCMD is indicated in lower case. The authentic stop codon is coloured red, and the new stop codon is coloured blue. **d, e**, Northern blot analysis of *fukutin* in human lymphoblasts (**d**) and model mice (**e**); F, FCMD; N, normal control. The wild-type mouse *fukutin* mRNA was detected at a size of 6.1 kb. Both skeletal muscle (left) and brain (right) showed smaller, abnormal bands in Hp/Hp mice. WT, wild type; Hn, Hn/Hn mice; Hp, Hp/Hp mice. **f**, Representation of genomic DNA and cDNA in ARH (*LDLRAP1*, left), NLSMD (*PNPLA2*, middle) and human (*AB627340*, right).

normal and aberrant forms of the protein in human and mouse using immunoprecipitation followed by western blot analysis. The abnormal *fukutin* protein in FCMD displayed the predicted mobility shift (Fig. 2a–c and Supplementary Fig. 6b).

We introduced normal and aberrantly spliced *fukutin* cDNA constructs into mammalian cell lines. Whereas normal *fukutin* localized to the Golgi apparatus, the aberrantly spliced *fukutin* protein is displaced completely from the Golgi to the endoplasmic reticulum (Fig. 2d and Supplementary Fig. 8). Further examination showed that a *fukutin* construct lacking the C-terminal 38 amino acids also mislocalized to the endoplasmic reticulum (Fig. 2d and Supplementary Fig. 8), suggesting that the C-terminal domain of *fukutin* is important for localization to the Golgi. Thus, impairment of this domain may lead to *fukutin* dysfunction in FCMD. The mislocalization is unlikely to be toxic because FCMD is an autosomal recessive disease and heterozygous carriers of the SVA insertion have no symptoms.

We next tested if exon-trapping occurs in other diseases with SVA insertion⁶. In a patient with autosomal recessive hypercholesterolemia (ARH), a 2.6-kb SVA was inserted within intron 1 of the *LDLRAP1* gene⁴. A patient with lipid storage disease with subclinical myopathy (NLSMD) also had a 1.9-kb SVA insertion in exon 3 of the *PNPLA2* gene⁵. We found abnormally spliced products induced by SVA exon-trapping in these patients' fibroblast (Fig. 1f left and middle panels, Supplementary Figs 9 and 10, and Supplementary Table 1). Cycloheximide treatment to fibroblasts from these patients increased expression of the genes (Supplementary Figs 9a and 10a), suggesting

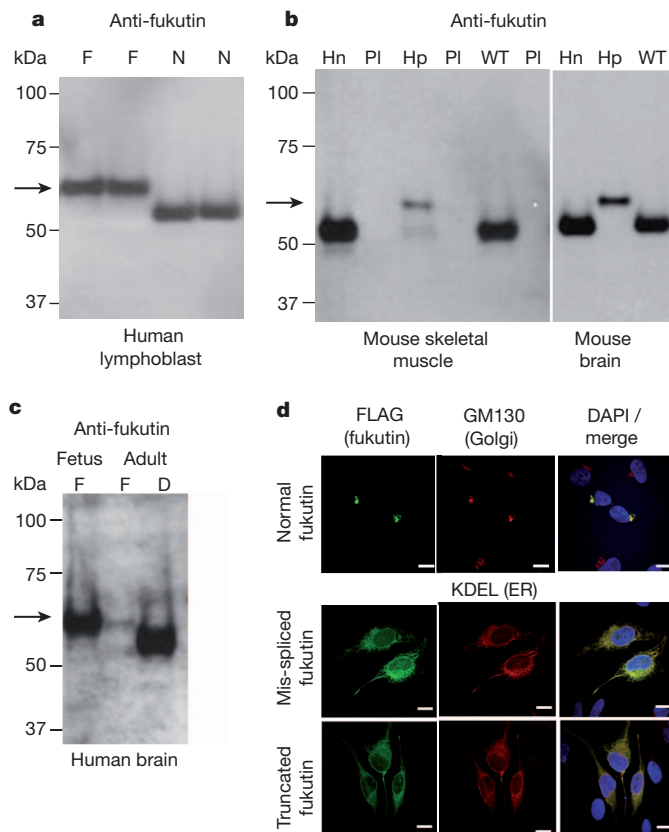


Figure 2 | Abnormal fukutin protein in FCMD. **a–c**, Immunoprecipitation analysis of fukutin protein in human lymphoblasts (**a**), both skeletal muscle and brain tissues from Hp/HP mice (**b**) and brain tissue from patients with FCMD (**c**); filled arrow, abnormal fukutin; N, normal sample; F, sample from patient with FCMD; Hn, Hn/Hn mice; Hp, Hp/HP mice; PI, pre-immune serum; D, patient with Duchenne muscular dystrophy. **d**, The subcellular localization of fukutin. Top, normal fukutin; middle, mis-spliced fukutin; bottom, truncated fukutin. Stained with anti-FLAG (left, to detect fukutin), anti-GM130 (middle, Golgi marker, top) and anti-KDEL (endoplasmic reticulum marker, middle and bottom), and merge (right, with DAPI stain). Scale bar, 10 μ m.

that the SVA-trapped transcripts are likely to be subjected to non-sense-mediated mRNA decay^{6,17}. In a search for the same events using the same acceptor site as FCMD in the human genome, we located two expressed sequence tags on human chromosome 4 (DA436529 and DA060755) that represent a spliced transcript induced by an SVA element. We found exonization in a human-specific insertion of SVA (AB627340) into a small gene (Fig. 1f right panel and Supplementary Fig. 11). The human-specific exon-trapping of SVA in the small gene might influence human evolution and development.

FCMD alleles of the *fukutin* gene contain a fully intact protein coding sequence, raising the possibility that FCMD could be treated by restoring translation of the full-length protein through splicing modulation with AONs. To identify promising target sequences in various cell lines, we produced 25-mer 2'-O-methyl phosphoramidite (2'OMePS) AONs targeted to the acceptor (A1–A3), donor (D1–D5) and exonic splicing enhancer sites (E1–E4) in *fukutin* pre-mRNA (Supplementary Fig. 12). We introduced the AONs into various cell types and assessed the recovery of normal processing and restoration of the authentic stop codon (Fig. 3a). Cells with A3 and E3 showed strong suppression of SVA-derived splicing. The greatest recovery of *fukutin* mRNA, to levels of more than 40% of the normal control, was achieved with a combination of A3, E3 and D5 (AED) (Fig. 3a). The D5 sequence overlaps with a predicted intronic splicing enhancer site within the aberrant intronic sequence; in normal *fukutin*, this sequence resides in exon 10 (Supplementary Fig. 12).

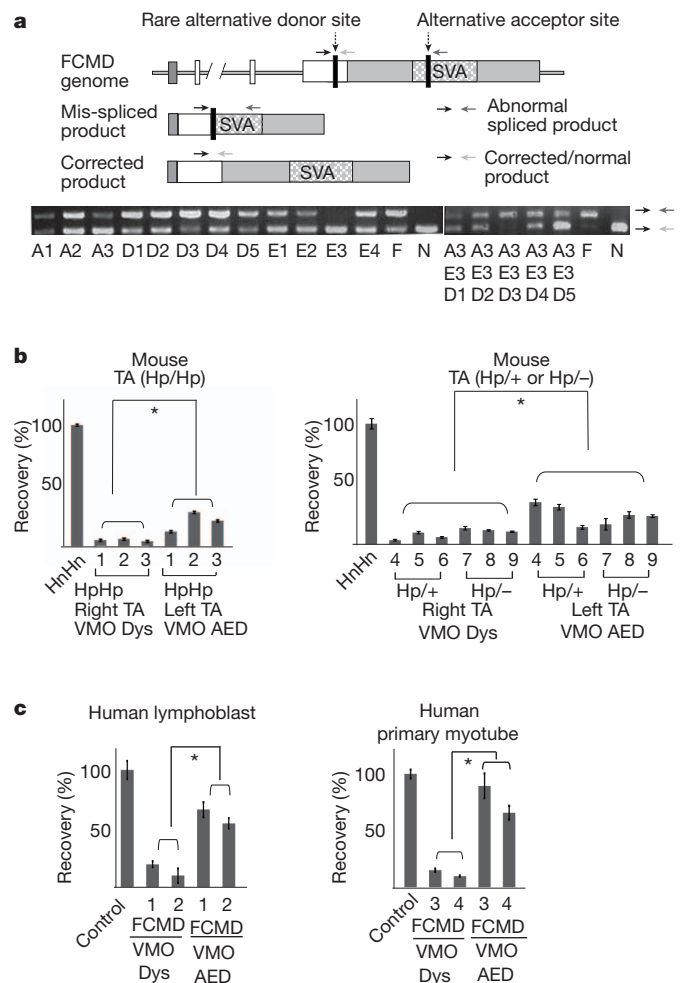


Figure 3 | AON cocktail rescues normal *fukutin* mRNA. **a**, RT-PCR diagram of three primers designed to assess normal *fukutin* mRNA recovery (upper). Black arrow, a common forward primer located on *fukutin* coding region; dark grey arrow, a reverse primer to detect the abnormal RT-PCR product (161 bp); light grey arrow, the other reverse primer to detect the restored normal RT-PCR product (129 bp). The effect on Hp/HP ES cells treated with each single or a cocktail of AONs (lower). F, FCMD; N, normal sample. **b**, Rescue from abnormal splicing in VMO-treated Hp/HP and Hp/– mice. Local injection of AED cocktail into tibialis anterior ($n = 3$). Dys, a negative control. **c**, Rescue from abnormal splicing in VMO-treated human FCMD lymphoblasts (left, $n = 2$) and myotubes (right, $n = 2$). The y axis shows the percentage recovery of normal mRNA (* $P < 0.01$ by Student's *t*-test). TA, tibialis anterior. Error bars, s.e.m.

We injected octa-guanidine morpholino oligonucleotide (vivomorpholino, VMO)¹⁸ AED cocktail locally into skeletal muscle of knock-in mice and evaluated the therapeutic effect by calculating the percentage recovery of normally processed mRNA. In the AED-treated tibialis anterior and gastrocnemius of Hp/HP and Hp/– mice, the amount of corrected *fukutin* mRNA increased significantly relative to mice treated with control VMO (Fig. 3b and Supplementary Fig. 13). We assessed fukutin protein recovery in injected skeletal muscle tissue from Hp/HP mice. Consistent with the significant increase of restored normal mRNA, normal fukutin protein was rescued (Fig. 4a). We examined α -DG glycosylation in AED-treated Hp/– mice. Deficiently glycosylated α -DG, at the predicted smaller size, was reduced in abundance, whereas normal-sized α -DG increased after AED treatment (Fig. 4b). The signal intensity for glycosylated α -DG was clearly increased, and a shift in the α -DG core was observed, indicating that the rescued fukutin is functional. Laminin overlay assays revealed a marked increase in α -DG laminin-binding ability, indicating that α -DG

15. Hancks, D. C. *et al.* Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum. Mol. Genet.* **20**, 3386–3400 (2011).
16. Kanagawa, M. *et al.* Residual laminin-binding activity and enhanced dystroglycan glycosylation by LARGE in novel model mice to dystroglycanopathy. *Hum. Mol. Genet.* **18**, 621–631 (2009).
17. Hancks, D. C. & Kazazian, H. H. Jr. SVA retrotransposons: evolution and genetic instability. *Semin. Cancer Biol.* **20**, 234–245 (2010).
18. Wu, B. *et al.* Octa-guanidine morpholino restores dystrophin expression in cardiac and skeletal muscles and ameliorates pathology in dystrophic mdx mice. *Mol. Ther.* **17**, 864–871 (2009).
19. Barresi, R. *et al.* LARGE can functionally bypass α -dystroglycan glycosylation defects in distinct congenital muscular dystrophies. *Nature Med.* **10**, 696–703 (2004).
20. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
21. Kazazian, H. H. Jr. Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
22. Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature Rev. Genet.* **10**, 691–703 (2009).
23. Hassoun, H. *et al.* A novel mobile element inserted in the α spectrin gene: spectrin dayton. A truncated α spectrin associated with hereditary elliptocytosis. *J. Clin. Invest.* **94**, 643–648 (1994).
24. Rohrer, J. *et al.* Unusual mutations in Btk: an insertion, a duplication, an inversion, and four large deletions. *Clin. Immunol.* **90**, 28–37 (1999).
25. Legoux, P. *et al.* Molecular characterization of germline NF2 gene rearrangements. *Genomics* **65**, 62–66 (2000).
26. Makino, S. *et al.* Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *Am. J. Hum. Genet.* **80**, 393–406 (2007).
27. O'Brien, S. *et al.* Randomized phase III trial of fludarabine plus cyclophosphamide with or without oblimersen sodium (Bcl-2 antisense) in patients with relapsed or refractory chronic lymphocytic leukemia. *J. Clin. Oncol.* **25**, 1114–1120 (2007).
28. Crooke, S. T. *et al.* Vitravene—another piece in the mosaic. *Antisense Nucleic Acid Drug Dev.* **8**, vii–viii (1998).
29. Lu, Q. L. *et al.* Functional amounts of dystrophin produced by skipping the mutated exon in the mdx dystrophic mouse. *Nature Med.* **9**, 1009–1014 (2003).
30. Alter, J. *et al.* Systemic delivery of morpholino oligonucleotide restores dystrophin expression bodywide and improves dystrophic pathology. *Nature Med.* **12**, 175–177 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Nakagawa, K. Ohno, S. Tsujino, N. Taniguchi, and I. Nonaka for comments; M. Okabe and A. Kawai for generating the ES cell line from knock-in model mice; Y. Motoyoshi and J. C. Cohen for providing patients' samples; W. Sako and Y. Izumi for sending patients' samples; I. Mizuta, T. Mure, M. Furukawa, K. Kaneshiro, Y. Dainin and all laboratory members for technical support; and J. Logan for editing the manuscript. We thank the GAIN for providing chimpanzee brain samples. This work was supported by an Intramural Research Grant (20B-13) for Neurological and Psychiatric Disorders from the National Center of Neurology and Psychiatry (to T.T.), the Global COE Program (Frontier Biomedical Science Underlying Organelle Network Biology) (to T.T., M.T.-I. and M.K.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, Grants-in-Aid for Scientific Research (A) (23249049 to T.T.), and Young Scientists (A) (21689030 to K.K.) and (B) (20790980 to M.T.-I.) from the Japan Society for the Promotion of Science, and the Takeda Science Foundation (to K.K.).

Author Contributions M.T.-I., K.K., M.K. and T.T. designed the study. M.T.-I. performed most of the experiments. K.K. developed a system to detect endogenous fukutin protein. M.K. performed biochemical analysis of VMO-injected mice. C.Y. produced the *fukutin* cDNA constructs for transfection experiments. K.M., T.O., and A.K. performed analyses of AON treatment in mice and various cell types. H.K., T.Y. and S.T. provided intellectual input. H.O.A., S.D. and R.K., provided patients' samples. M.T.-I., K.K. and T.T. wrote the paper.

Author Information The patient *fukutin* and a chimpanzee mRNA sequences are deposited in GenBank/European Molecular Biology Laboratory/DNA Data Bank of Japan under accession numbers AB609007 and AB627340, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.T. (toda@med.kobe-u.ac.jp).

METHODS

Antisense oligonucleotides. Twenty-five-mer 2'OMePS (GeneDesign and Invitrogen) and VMO oligonucleotides (Gene-Tools) were designed to target potential splice-modulating sequences of SVA-inserted *fukutin*, including a splicing acceptor site, a splicing donor site, exonic splicing enhancers and intronic splicing enhancers as follows: A1, CCGTGGAAGGAGACTGTGGAGGGAG; A2, GGAGACCGTGAAGGAGACTGTGGA; A3, AGAGGGAGACCGTGAAGGAGACTG; E1, CACCGTCCAGCCTTGGCTCGGCATC; E2, CTGCAGTGAGCCGAGATGGCAGCAG; E3, GAGGCAGGAGAATCAGGCAGGGAGG; E4, GAAAACCACTGAGGCGTAGCAGGCT; D1, CAGGTCTTACCATAGTGGCTTCAA; D2, CAGGAATCTTCCAGGTCTTACCATA; D3, GAGCGCTTCCAGTCCACGTCTTTA; D4, TCCATTGGGTTCACATTGGGAGGA; D5, CATCCCACTCAGAAATAGGCCAGAT; DYS, GGCCAAACCTCGGCTTACCTGAAAT³¹. U (uracil) was used instead of T (thymine) for the synthesis of 2'-O-MePS oligonucleotides. Target sequences are shown in Supplementary Fig. 12. Exonic splicing enhancer sites were predicted by ESEfinder 3.0 (<http://rulai.csh.edu/cgi-bin/tools/ESE3/ese finder.cgi>), and intronic splicing enhancer sites were predicted by ACESCAN2 (<http://genes.mit.edu/acescan2/index.html>). AONs were solubilized in sterile distilled water.

Animals and cells. All mouse experimental protocols were approved by the Ethics Review Committees for Animal Experimentation at Osaka University Graduate School of Medicine and Kobe University Graduate School of Medicine. FCMD knock-in model mice and the mouse nomenclature have been described previously¹⁶. The transgenic alleles containing normal and SVA-inserted human exon 10 were named Hn and Hp, respectively: Hp/Hp is homozygous for the SVA allele; Hn/Hn is homozygous for the normal allele; Hp/+ and Hp/- are SVA carriers and compound heterozygotes for the SVA and knockout alleles, respectively. The ages of mice used in experiments varied from 2 to 6 months. The mouse ES cell line carrying the SVA-inserted human genomic *fukutin* exon 10 was generated from Hp/Hp mice. The ES cell line carrying a *fukutin* knockout allele has been described previously³². The commercially available mouse ES cell line AB2.2 was used as a control. Human lymphoblasts were obtained from patients with FCMD with homozygous SVA insertions and from unaffected individuals. Human primary myoblasts were derived from muscle biopsies from patients with FCMD and unaffected individuals. Human primary fibroblasts were obtained by skin biopsy from patients with ARH and NLSMD. Human autopsy brain samples were obtained from patients with FCMD (fetus and 34-year-old) and DMD (34-year-old). Chimpanzee brain sample was provided by the Great Ape Information Network, Japan. Human brain RNA was purchased from Clontech. All clinical samples were used with the approval of Human Ethics Review Committees of Osaka University Graduate School of Medicine and Kobe University Graduate School of Medicine.

Myoblast differentiation. Myoblast cells were maintained at 37 °C and 5% CO₂ in DMEM medium plus 20% fetal bovine serum, 2.5 ng ml⁻¹ of basic fibroblast growth factor (Sigma), and a 0.5% penicillin-streptomycin-amphotericinB mix (Wako). Myotubes were obtained from confluent myoblast cultures after 10–14 days of serum deprivation and replacement with 2% FBS.

RNA isolation, RT-PCR, qRT-PCR and sequencing. To inhibit nonsense-mediated mRNA decay, cycloheximide (100 µg ml⁻¹) (Sigma) was added to the culture medium 24 h before RNA isolation. For RT-PCR and qRT-PCR, total RNA was extracted using the RNeasy Plus Mini kit (Qiagen), and cDNA was obtained using the Superscript III One-step RT-PCR system (Invitrogen) with random primers, following the manufacturer's instructions. SYBR Pre-mix Ex Taq (Takara) was used for qRT-PCR, and expression values were normalized to *gapdh* as an internal control for mRNA quantity. Data were obtained from triplicate experiments. To detect abnormally spliced RT-PCR products from patients with FCMD, ARH and NLSMD, and from human brain AB627340 cDNA, long-range PCR was performed using LA Taq with LA Taq Buffer II (Takara), adding dimethyl sulphoxide and 7-deaza-dGTP (Roche). The RT-PCR products were directly sequenced (FCMD and NLSMD), or cloned with the TOPO TA Cloning Kit (Invitrogen) before sequencing (ARH and AB627340). To calculate the expression ratio in Fig. 1a and Supplementary Figs 4, 5, 9, 10 and 13, the value in the mutant sample was divided by the value in the normal sample, as measured by qRT-PCR. To identify AON target sequences, we designed three primers to distinguish recovered transcripts from unrecovered transcripts by AON treatment (Fig. 3a). Similarly, we designed three primers to compare expression amount of SVA-trapped to SVA-untrapped transcripts of the AB627340 gene (Supplementary Fig. 11a). One primer on SVA in Fig. 3a and Supplementary Fig. 11a was within *Alu*-like domain: the sequence was 5'-GAAAACCACTGAGGCGTAGC-3'. To calculate the percentage recovery of normal mRNA processing in Fig. 3b, c and Supplementary Fig. 13, the value of treated sample was divided by that of normal samples, as measured by qRT-PCR at sequence position 1341, where the authentic

stop codon resides. Primer sequences for qRT-PCR and RT-PCR are available upon request.

Northern blot analysis. Previous attempts to detect *fukutin* mRNA in patients with FCMD by northern blot analysis have been unsuccessful³, probably because the predicted mRNA sequence is the same size as abundant ribosomal RNA. Moreover, the tertiary structure of *fukutin* mRNA is presumably complicated owing to the immensely GC-rich SVA sequence. Therefore, we performed northern blot analysis of FCMD and control mRNA after treatment to remove abundant ribosomal RNA and strong denaturation to untangle the *fukutin* transcript. Total RNA (1 mg) was extracted from human lymphoblasts, mouse ES cells, mouse brain and mouse skeletal muscle using TRIzol (Invitrogen). Oligotex-dT30<Super> (Takara) was used to extract more than 3 µg of poly-A RNA. Ribosomal RNA was removed using Ribo-Minus (Invitrogen). Stronger denaturation of RNA was achieved by incubating poly-A-RNA samples with a combination of 0.8 M glyoxal and 50% DMSO in 10 mM sodium phosphate buffer (pH 7.0) for 60 min at 55 °C. Three micrograms of poly-A RNA was loaded on the agarose gel. A *fukutin* cDNA clone covering the *fukutin* coding sequence was ³²P-labelled and used as a probe.

cDNA expression constructs. The normal *fukutin* cDNA encodes full-length *fukutin* protein. The spliced *fukutin* construct encodes abnormal *fukutin*, as shown in Supplementary Fig. 7. The truncated *fukutin* construct lacks the C-terminal 38 amino acids. All constructs encoded FLAG epitope tags fused to the C terminus of the expressed protein.

Cell transfection. HeLa S3 cells and C2C12 cells were transfected with normal *fukutin* construct, spliced *fukutin* construct and truncated *fukutin* construct using FuGENE 6 (Roche). *Fukutin* localization was determined using immunocytochemistry 2 days after transfection. For transfection of AONs, 2'OMePS were introduced into various cell lines, including mouse ES cells, human myoblasts and human lymphoblasts, using Lipofectin (Invitrogen).

Detection of endogenous *fukutin* protein. The polyclonal rabbit anti-*fukutin* antibody RY213 recognizes the peptide CLKIESKDPRLDGIDS, and the polyclonal goat-anti-*fukutin* antibody 106G2 recognizes full-length *fukutin* protein lacking the amino (N)-terminal hydrophobic domain. Endogenous *fukutin* was detected by immunoprecipitation using 106G2, from cell or tissue lysates containing 5–10 mg of total protein in lysis buffer (1% Nonidet P-40, 0.5% deoxycholate, 0.1% SDS, 20 mM Tris-Cl, pH 7.5 and 150 mM NaCl), followed by western blot analysis using affinity-purified RY213.

Immunofluorescence and western blot analysis. Cells were washed and fixed with 4% paraformaldehyde in PBS. The following primary antibodies were used: anti-GM130 (monoclonal, BD Bioscience), anti-KDEL (monoclonal, Stressgen), anti-FLAG (rabbit polyclonal, MBL), anti-α-DG (monoclonal, I1H6C4 and VIA4-1, Millipore) and anti-laminin (rabbit polyclonal, Sigma). To stain nuclei, 4',6'-diamidino-2-phenylindole (DAPI, Sigma) was added to the secondary antibody solution at a final concentration of 1 ng ml⁻¹. Cells were observed under fluorescence confocal microscopy (Carl Zeiss). Western blot analysis and laminin overlay assays were performed as described previously¹⁶.

Mutagenesis analysis. We made the four *fukutin* constructs: pHn, human normal *fukutin* construct consisting of exon 2–9 cDNA and genomic normal exon 10; pHp, patient *fukutin* construct consisting of exon 2–9 cDNA and genomic patient exon 10 with SVA insertion; pSpl, patient *fukutin* construct pHp, which lacks the abnormally spliced region; pAcc, patient *fukutin* construct pHp with AG to GG replacement at the acceptor site within the SVA sequence. These constructs were transfected into HeLa S3 cells using Effectene (Qiagen). After extraction of poly-A RNA by Oligotex, northern blot analysis was performed using 2 µg of poly-A RNA for each sample with stronger denaturation mentioned above.

AON treatment of FCMD model mice. For intramuscular injection, we injected cardiotoxin (10 µM) (Latoxan) percutaneously into tibialis anterior (0.3 nmol) and gastrocnemius (0.7 nmol) of Hp/+, Hp/–, Hp/Hp and Hn/Hn mice on day 0 (*n* = 3 for each genotype). On days 1, 4 and 7, VMO (400 mg kg⁻¹) solubilized in sterile distilled water was injected. AED and Dys were administered to the left and the right legs, respectively. For systemic injection, an intraperitoneal injection of butorphanol tartrate (5 mg kg⁻¹) (Bristol-Myers Squibb) was performed on day 0. VMO (20 mg kg⁻¹) solubilized in 5% glucose solution was administered by intravenous injection through the tail vein on days 1 and 7 (*n* = 4 for Hp/–, *n* = 2 for Hp/+). Mice were killed on day 21, and total RNA or protein lysate was isolated from each tissue for further analyses of *fukutin* mRNA expression, *fukutin* protein translation, and glycosylation of α-DG.

AON treatment of human patient cell lines. For protein analysis, VMO cocktails (AED and Dys) were introduced into FCMD and normal control lymphoblasts at a final concentration of 2.5 µM in culture medium using a Gene Pulser II Electroporator (0.25-kV voltage, 960-µF capacitance, with 0.4-cm gene pulser cuvettes, giving a time-constant readout of approximately 40 ms) (Bio-Rad) (*n* = 2). For glycosylation analysis, VMO cocktails (AED and Dys) were introduced into myoblasts from patients with FCMD and normal control cells by direct

addition to the culture medium at a final concentration of 4 μ M ($n = 2$). After incubation for 48 h, cells were collected and total RNA or protein lysate was isolated. **Laminin clustering assay.** The AED cocktail was introduced into myotubes by direct addition to the culture medium at a total concentration of 4 μ M after a medium change on day 2. On days 10–14, mouse EHS laminin-1 (Sigma) was added with fresh medium at a concentration of 1.0 nM and incubated for 30 min, followed by immunocytochemistry.

SVA sequence analysis. SVA sequence was aligned to the SVA reference sequence present in Repbase (<http://www.girinst.org/repbase/update/index.html>)³³ and the

location on the SVA reference of the splicing acceptor and donor sites in SVA was determined.

31. Yokota, T. *et al.* Efficacy of systemic morpholino exon-skipping in Duchenne dystrophy dogs. *Ann. Neurol.* **65**, 667–676 (2009).
32. Takeda, S. *et al.* Fukutin is required for maintenance of muscle integrity, cortical histogenesis and normal eye development. *Hum. Mol. Genet.* **12**, 1449–1459 (2003).
33. Jurka, J. Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet.* **9**, 418–420 (2000).

ATP-induced helicase slippage reveals highly coordinated subunits

Bo Sun^{1,2*}, Daniel S. Johnson^{1,2,†*}, Gayatri Patel³, Benjamin Y. Smith^{1,2}, Manjula Pandey³, Smita S. Patel³ & Michelle D. Wang^{1,2}

Helicases are vital enzymes that carry out strand separation of duplex nucleic acids during replication, repair and recombination^{1,2}. Bacteriophage T7 gene product 4 is a model hexameric helicase that has been observed to use dTTP, but not ATP, to unwind double-stranded (ds)DNA as it translocates from 5' to 3' along single-stranded (ss)DNA^{2–6}. Whether and how different subunits of the helicase coordinate their chemo-mechanical activities and DNA binding during translocation is still under debate^{1,7}. Here we address this question using a single-molecule approach to monitor helicase unwinding. We found that T7 helicase does in fact unwind dsDNA in the presence of ATP and that the unwinding rate is even faster than that with dTTP. However, unwinding traces showed a remarkable sawtooth pattern where processive unwinding was repeatedly interrupted by sudden slippage events, ultimately preventing unwinding over a substantial distance. This behaviour was not observed with dTTP alone and was greatly reduced when ATP solution was supplemented with a small amount of dTTP. These findings presented an opportunity to use nucleotide mixtures to investigate helicase subunit coordination. We found that T7 helicase binds and hydrolyses ATP and dTTP by competitive kinetics such that the unwinding rate is dictated simply by their respective maximum rates V_{\max} , Michaelis constants K_M and concentrations. In contrast, processivity does not follow a simple competitive behaviour and shows a cooperative dependence on nucleotide concentrations. This does not agree with an uncoordinated mechanism where each subunit functions independently, but supports a model where nearly all subunits coordinate their chemo-mechanical activities and DNA binding. Our data indicate that only one subunit at a time can accept a nucleotide while other subunits are nucleotide-ligated and thus they interact with the DNA to ensure processivity. Such subunit coordination may be general to many ring-shaped helicases and reveals a potential mechanism for regulation of DNA unwinding during replication.

Despite the fact that most motor proteins use ATP as a fuel source, previous bulk studies have shown that T7 helicase does not unwind DNA efficiently in the presence of ATP, although it is capable of ATP hydrolysis^{5,6,8}. To investigate why ATP seemed not to support T7 helicase unwinding, we used a single-molecule optical trapping assay that we previously developed to measure unwinding of dsDNA or translocation on ssDNA (Fig. 1a and Supplementary Fig. 1)⁹. Briefly, two strands of a DNA fork junction were held under tension that was not sufficient to mechanically unwind the junction without a helicase. Helicase unwinding of the junction resulted in an increase in the ssDNA length, permitting tracking of the helicase location. When experiments were conducted with 2 mM ATP, we were surprised to find that ATP supported not only dsDNA unwinding but that it also supported it at a significantly faster rate than with dTTP (Fig. 1b–c). However, processive unwinding was interrupted by slippage events, resulting in a remarkable sawtooth pattern in the unwinding trace

(Fig. 1b). Control experiments verified that each trace was the action of a single helicase (Supplementary Fig. 2). We attribute this pattern to helicase losing its grip on the ssDNA, sliding backwards under the influence of the reannealing DNA fork, and then regaining its grip and resuming unwinding (Fig. 1d). In contrast, slippage behaviour was essentially absent with 2 mM dTTP alone (Fig. 1b). These results resolve the mystery of the apparent lack of significant unwinding activity seen in bulk studies^{4–6,8}; unwinding and slippage could not be separated, so unwinding was masked by unobservable slips that prevented helicase from moving over a substantial distance. Our work is the first direct observation, to our knowledge, of helicase nucleotide-specific slippage. Previous studies of non-ring-shaped helicases have reported reverse motions of the unwinding fork attributable to helicase reaching the end of the DNA or encountering a barrier^{10,11}, dissociating from the DNA^{12,13}, or moving in the reverse direction^{9,12,13}. These are of a somewhat different nature than what we have observed. The only slippage behaviour that may resemble ours is from non-helicase bacteriophage motors^{14,15}, but their slippage is not a result of the use of a specific nucleotide.

Slippage was not observed with dTTP alone (Fig. 1b) and therefore seems to be sensitive either to the base composition of the bound nucleotide (for example, adenosine versus thymidine) or the type of sugar (ribose versus deoxyribose). We compared slippage for all four NTPs and their dNTP counterparts (Supplementary Fig. 3). For each nucleotide we measured processivity, defined as the mean distance between slips (Supplementary Fig. 4). The results indicate that the additional 2'-OH group on the ribose sugar makes the helicase more prone to slipping. Examination of the helicase structure at the nucleotide-binding pocket¹⁶ reveals that the 2'-OH group of a bound nucleotide may displace the -OH group on the side chain of residue Y535 (Supplementary Fig. 5a). We thus generated a Y535F mutant to remove the -OH group and it showed significantly increased processivity in the presence of ATP, albeit still less than that seen for dATP (Supplementary Fig. 5b).

Although ATP caused helicase to slip more frequently, it supported a much faster unwinding rate between slips, consistent with an earlier finding of a faster rate of ATP hydrolysis¹⁷. Because ATP and dTTP support different unwinding rates and processivities, we used nucleotide mixtures to understand how multiple subunits of the helicase coordinate unwinding activity. We approximated the *in vivo* concentrations of ATP and dTTP of *Escherichia coli*¹⁸ by using 2.0 mM ATP and a small amount of dTTP, 0.2 mM (Fig. 1b, c). Although the unwinding rate between slips was close to the value observed with 2 mM ATP alone, the processivity increased by approximately threefold. When the converse experiment was performed (0.2 mM ATP and 2.0 mM dTTP), the unwinding rate was comparable to that with 2 mM dTTP alone and minimal slippage was observed (Fig. 1b, c). These results imply that even a small fraction of helicase subunits, when bound with dTTP, reduce slippage and substantially increase processivity. This finding was further substantiated by bulk experiments using ATP alone, and an ATP/dTTP mixture (Supplementary Fig. 6). To determine if T7

¹Department of Physics - Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA. ²Howard Hughes Medical Institute, Cornell University, Ithaca, New York 14853, USA.

³Department of Biochemistry, UMDNJ-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, USA. [†]Present address: Rockefeller University, New York, New York 10065, USA.

*These authors contributed equally to this work.

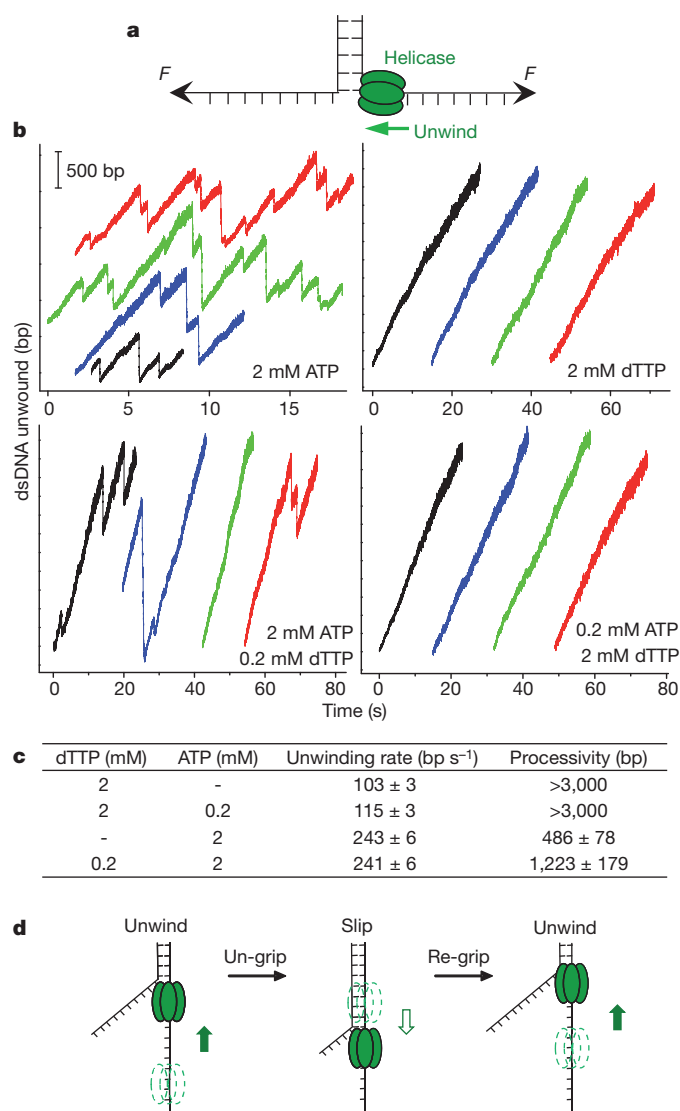


Figure 1 | Comparison of helicase unwinding behaviours with different nucleotides. **a**, Schematic of the single-molecule configuration (not to scale). The single-stranded ends of a dsDNA were held at a constant unzipping force of 8 pN while T7 helicase unwound the dsDNA by translocating on ssDNA. **b**, Representative traces showing the number of unwound base pairs versus time in the presence of various concentrations of nucleotides. For clarity, traces have been arbitrarily shifted along both axes. **c**, A summary of unwinding rates and processivities. Uncertainties are s.e.m. **d**, Cartoon illustrating slippage behaviour. The helicase unwinds, loses grip, slips, re-grips and resumes unwinding. Dotted helicase indicates a previous location of the helicase.

helicase binds DNA with different affinities in the presence of dTTP and ATP, bulk binding studies were carried out using fluorescence anisotropy with dTTP and ATP analogues (Supplementary Fig. 7). The results show that T7 helicase binds ssDNA 100-fold more tightly with dTMPCCP than with AMPPCP, and indicate that the greater slippage in the presence of ATP is probably due to weaker binding to DNA.

The discovery of helicase slippage and the ability to directly measure helicase processivity provided a unique opportunity to explore the following: (1) how ATP and dTTP compete for binding to helicase subunits; (2) how nucleotide binding regulates helicase affinity to DNA; and (3) how multiple subunits of helicase coordinate their activities.

To understand how ATP and dTTP compete for binding to helicase subunits, we determined the unwinding rates between slippage events (Fig. 2a) as a function of nucleotide concentration. For each nucleotide alone, the unwinding rate followed Michaelis–Menten-like kinetics, yielding V_{\max} and K_M values that were both higher for ATP than for

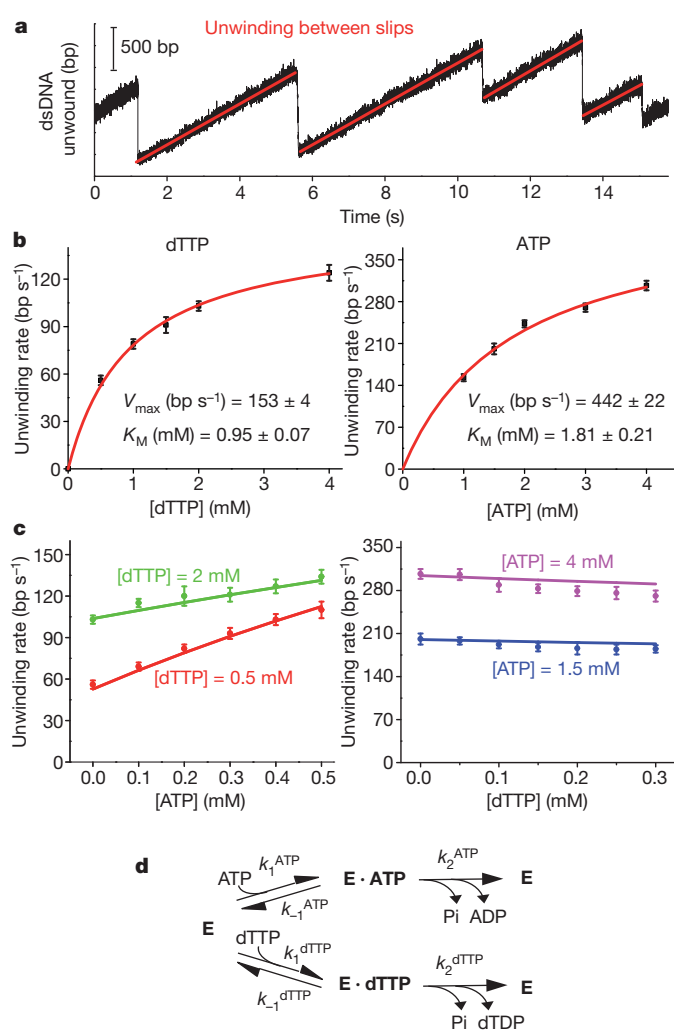


Figure 2 | Helicase unwinding kinetics. **a**, Example of unwinding with ATP to illustrate the method of determining unwinding rate by analysing data between slips. **b**, Kinetic constants for unwinding under a constant unzipping tension of 8 pN in the presence of either ATP (right) or dTTP (left). For each nucleotide, K_M and V_{\max} were obtained by fitting the unwinding rates as a function of NTP concentration to the Michaelis–Menten equation. **c**, Measured unwinding rates at either fixed [dTTP] and varying [ATP], or fixed [ATP] and varying [dTTP], and comparison with direct predictions (not fits) from the competitive nucleotide binding model using kinetic constants K_M and V_{\max} shown in **b**. Error bars indicate s.e.m. **d**, Kinetic pathway of a competitive binding model where ATP and dTTP compete for binding and hydrolysis by the helicase (denoted by E here).

dTTP (Fig. 2b). These kinetics indicated that there was no cooperativity in NTP binding and hydrolysis. Next, we conducted experiments in which the concentration of one nucleotide was fixed while that of the other nucleotide was varied. The resulting unwinding rates could be explained by competitive kinetics: ATP and dTTP compete for binding based on their respective affinities and the resulting reaction rate is determined by their concentrations, V_{\max} and K_M (Fig. 2c, d; Methods Summary and Supplementary Discussion). A comparison of unwinding rates with mixed nucleotides and direct predictions (not fits) from the competitive binding kinetics showed excellent agreement. These results were further substantiated by ssDNA translocation rate experiments (Supplementary Fig. 8). This also explains why in Fig. 1b, c the unwinding rate was minimally altered when 0.2 mM of dTTP was added to 2 mM ATP. Under those conditions, only about 16% of the nucleotide bound to the helicase hexamer was dTTP.

The competitive binding kinetics for nucleotides, however, does not explain the observed slippage behaviour with mixed nucleotides (Fig. 1b, c). That is, it is unclear how the 16% bound dTTP resulted

in a threefold increase in processivity. If only a single nucleotide can be bound by the helicase at a time and the type of the bound nucleotide determines the helicase's affinity to the DNA, then processivity should only increase by 7% (Supplementary Discussion). In addition, it has previously been shown that the helicase subunits do not bind to ssDNA in the absence of a nucleotide¹⁹. However, we found minimal slippage even at [dTTP] much below its K_M . These observations indicate participation of multiple subunits in both nucleotide and DNA binding, where each subunit would have a nucleotide-specific DNA binding affinity. Our data indicate that helicase may not slip if at least one subunit of the hexamer is in a deoxythymidine-ligated state, which has a higher affinity for the DNA.

Two models may be consistent with this idea. In an uncoordinated model^{1,2,7}, each helicase subunit functions independently in its nucleotide binding/hydrolysis, and DNA binding/release (Supplementary Discussion). Conversely, coordinated models have been proposed for T7 helicase^{1,2,7}, but details of the coordination remain unclear. Biochemical and structural studies indicate that nucleotide hydrolysis may occur sequentially around the hexameric ring^{16,20,21}, that roughly four subunits are nucleotide-ligated at any given time²⁰, and that DNA binding to the helicase might involve one-to-two helicase subunits^{16,20–22}. A model based on structural studies has been proposed for ring-shaped helicases E1 (ref. 23) and Rho²⁴, where all or some of the subunits coordinate their chemo-mechanical activities (Fig. 3d). Coordination

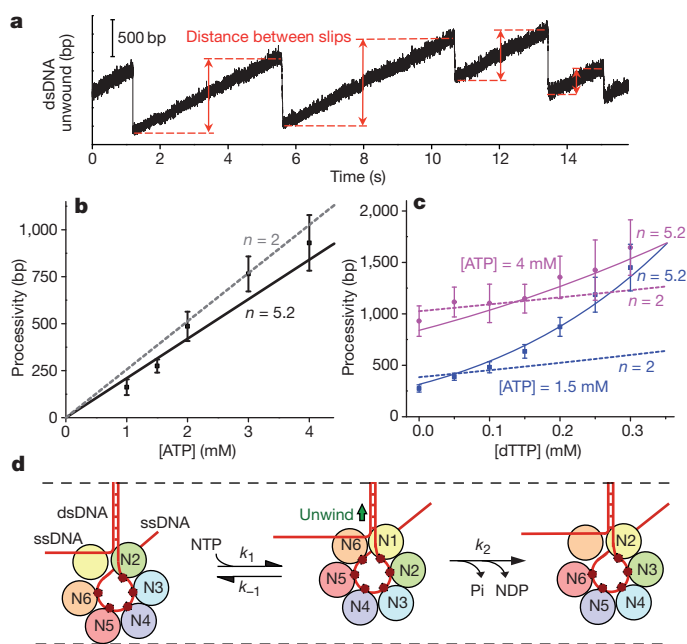


Figure 3 | Processivity dependence on nucleotides and a proposed coordinated model. **a**, An example of unwinding with ATP to illustrate the method of determining distance between slips. **b**, **c**, Measured processivity (mean distance between slipping events) as a function of [ATP] alone, and as functions of [dTTP] at two fixed concentrations of ATP. Note processivity increased substantially when a small amount of dTTP was added to the reaction. Solid lines are global fits using the coordinated model, yielding $n = 5.2 \pm 0.4$. For comparison, fits using $n = 2$ are also shown. Error bars indicate s.e.m. **d**, An interpretation of the proposed coordinated model. Each subunit is uniquely labelled with a different colour and has a potential ssDNA-binding site (small red dot). Nucleotide binding and subsequent hydrolysis occur sequentially around the ring. If a subunit is nucleotide-ligated (the state of hydrolysis indicated by Ni), it has a non-zero probability of being bound to ssDNA. During unwinding, the leading subunit can bind to a nucleotide (N) and thus acquire affinity to the upstream ssDNA. This stimulates the last nucleotide-bound subunit to release its nucleotide and ssDNA. Then the cycle proceeds again around the ring. Slippage occurs when all subunits simultaneously release ssDNA, as determined by the joint probability of detachment for all subunits (Supplementary Discussion).

could occur sequentially around the hexameric ring with the leading subunit poised for NTP binding and each successive subunit having a bound nucleotide in states of progression along the chemical reaction pathway (NTP, NDP + Pi, NDP, and so on). Depending on the state and type of nucleotide bound each subunit may have a different affinity to DNA. Once the leading subunit binds to an NTP and reels in the DNA, the remaining subunits progress to their next reaction states. Product release by the last participating subunit results in release of DNA from that subunit, and thus completes a single cycle.

We formulated quantitative descriptions for the uncoordinated and coordinated models (Supplementary Discussion). The observed rate of unwinding as a function of [ATP] or [dTTP] is consistent with both models, which predict an apparent Michaelis–Menten-like kinetics. The observed unwinding rate with ATP and dTTP mixtures is also consistent with the competitive binding kinetics for both models as long as, in the case of the coordinated model, the rates are treated as averages over time (Supplementary Discussion). Although the two models cannot be distinguished based on rate measurement studies, they do yield different predictions for DNA slippage behaviour. The uncoordinated model (Supplementary Discussion) requires that each subunit binds and hydrolyses nucleotides independently with an affinity to DNA dependent on the state and type of nucleotide bound. This model is not consistent with the processivity data taken with mixed nucleotides at concentrations near or lower than their respective K_M values (Supplementary Fig. 9).

On the other hand, the coordinated model requires that subunits participating in coordination bind and hydrolyse nucleotide in coordination, with only one subunit poised to bind a nucleotide at a time and with each subunit having an affinity to DNA dependent on the state and type of nucleotide bound. This model predicts that processivity should increase linearly with [NTP] in the presence of a single type of NTP. Indeed, our data show that the processivity increases linearly with increasing [ATP] (Fig. 3a, b). If multiple helicase subunits coordinate in their chemo-mechanical activities, what is the degree of coordination as measured by the number of participating subunits at any given time (n)? This is a key parameter that characterizes the mechanism of the helicase. Previous studies indicate that only one or two subunits are involved in significant DNA binding, suggesting a lower degree of coordination of $n = 1$ or 2 (refs. 16, 20–22). However, subunits may participate in the coordination even if they have lower affinity to ssDNA. The coordinated model formulated (Supplementary Discussion) is rather general and naturally takes this into account. Interestingly, it predicts that processivity sensitively depends on n as [dTTP] is increased in the presence of a fixed [ATP]—the larger n , the more subunits participate in DNA binding, and the more steeply processivity increases with [dTTP]. Therefore we measured processivity with mixtures of ATP and dTTP (Fig. 3c). A global fit to the processivity data in Fig. 3b, c yielded $n = 5.2 \pm 0.4$ (Methods Summary). In contrast, $n = 2$ does not agree with the measurements. These findings are further substantiated by experiments using UTP instead of ATP (Supplementary Fig. 10, $n = 5.0 \pm 0.3$), experiments under a different unzipping force (Supplementary Fig. 11, $n = 5.4 \pm 0.3$), and data on time between slips (Supplementary Fig. 12, $n = 5.5 \pm 0.4$). Because $n \leq 6$ is expected for a hexamer, this finding indicates that nearly all subunits participate in the coordination ($n = 5$ or 6) (Fig. 3d). Our findings suggest that only one subunit at a time can accept an incoming nucleotide, while the rest of the subunits are already nucleotide bound and coordinate to prevent slippage and maintain high processivity.

The work presented here provides a quantitative description of nucleotide binding/hydrolysis and its coupling to DNA binding and translocation for T7 helicase. This was possible because unwinding and slippage events are clearly distinguishable in single-molecule traces. The slippage behaviour is explained by a multiple-site coordinated model. For helicase to slip, all six subunits must simultaneously lose their grip on the DNA. This happens more often when helicase subunits are bound only to ribose nucleotides. Our data demonstrate

that T7 helicase has a very weak DNA binding affinity in the presence of ATP but the addition of a small amount of dTTP to the ATP reaction increases the binding affinity of helicase to DNA. As a consequence, the presence of a single deoxythymidine-ligated subunit significantly decreases the chance of slippage so that helicase can still effectively unwind dsDNA with ATP. Thus T7 helicase, like most other helicases², could still use ATP as a main power source *in vivo*, under conditions such as those during phage infection of *E. coli*¹⁸ where ATP is most abundant. ATP could be used for rapid unwinding and dTTP for high processivity. Although we focus here on a comparison of dTTP with ATP for helicase unwinding, other deoxyribose nucleotides may also reduce the frequency of slippage (Supplementary Fig. 3). We speculate that slippage may also provide an evolutionary advantage for replication: when dNTP concentrations are low, slippage can slow down helicase to allow its synchronization with a slow-moving DNA polymerase.

METHODS SUMMARY

Single-molecule assays were performed as described previously⁹. If dTTP and ATP compete for binding to helicase according to the kinetic pathway outlined in Fig. 2d, then the resulting unwinding rate is: $V_{\text{tot}} = \left(V_{\text{max}}^{\text{ATP}} \frac{[\text{ATP}]}{K_{\text{M}}^{\text{ATP}}} + V_{\text{max}}^{\text{dTTP}} \frac{[\text{dTTP}]}{K_{\text{M}}^{\text{dTTP}}} \right) / \left(1 + \frac{[\text{ATP}]}{K_{\text{M}}^{\text{ATP}}} + \frac{[\text{dTTP}]}{K_{\text{M}}^{\text{dTTP}}} \right)$, where for each type of nucleotide $K_{\text{M}} = \frac{k_{-1} + k_2}{k_1}$ and $V_{\text{max}} = sk_2$ with s being the step size (in nucleotides) (see Supplementary Discussion). In the presence of dTTP and ATP, if n helicase subunits coordinate in their chemo-mechanical activities and DNA binding, then the resulting distance between slips (processivity) is: $d_{\text{processivity}} = c \left(V_{\text{max}}^{\text{ATP}} \frac{[\text{ATP}]}{K_{\text{M}}^{\text{ATP}}} + V_{\text{max}}^{\text{dTTP}} \frac{[\text{dTTP}]}{K_{\text{M}}^{\text{dTTP}}} \right) / \left(\frac{[\text{ATP}]/K_{\text{M}}^{\text{ATP}}}{[\text{ATP}]/K_{\text{M}}^{\text{ATP}} + [\text{dTTP}]/K_{\text{M}}^{\text{dTTP}}} \right)^{n-1}$ (Supplementary Discussion), with c being a proportionality constant. This expression was used to fit data in Fig. 3b, c with c and n as fit parameters.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 December 2010; accepted 1 August 2011.

Published online 18 September; corrected 6 October 2011 (see full-text HTML version for details).

- Singleton, M. R., Dillingham, M. S. & Wigley, D. B. Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76**, 23–50 (2007).
- Patel, S. S. & Picha, K. M. Structure and function of hexameric helicases. *Annu. Rev. Biochem.* **69**, 651–697 (2000).
- Donmez, I. & Patel, S. S. Mechanisms of a ring shaped helicase. *Nucleic Acids Res.* **34**, 4216–4224 (2006).
- Matson, S. W., Tabor, S. & Richardson, C. C. The gene 4 protein of bacteriophage T7. Characterization of helicase activity. *J. Biol. Chem.* **258**, 14017–14024 (1983).
- Matson, S. W. & Richardson, C. C. DNA-dependent nucleoside 5'-triphosphatase activity of the gene 4 protein of bacteriophage T7. *J. Biol. Chem.* **258**, 14009–14016 (1983).
- Hingorani, M. M. & Patel, S. S. Cooperative interactions of nucleotide ligands are linked to oligomerization and DNA binding in bacteriophage T7 gene 4 helicases. *Biochemistry* **35**, 2218–2228 (1996).
- Lyubimov, A. Y., Strycharska, M. & Berger, J. M. The nuts and bolts of ring-translocase structure and mechanism. *Curr. Opin. Struct. Biol.* **21**, 240–248 (2011).
- Lee, S. J. & Richardson, C. C. Molecular basis for recognition of nucleoside triphosphate by gene 4 helicase of bacteriophage T7. *J. Biol. Chem.* **285**, 31462–31471 (2010).
- Johnson, D. S., Bai, L., Smith, B. Y., Patel, S. S. & Wang, M. D. Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. *Cell* **129**, 1299–1309 (2007).
- Myong, S., Rasnik, I., Joo, C., Lohman, T. M. & Ha, T. Repetitive shuttling of a motor protein on DNA. *Nature* **437**, 1321–1325 (2005).
- Myong, S., Bruno, M. M., Pyle, A. M. & Ha, T. Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. *Science* **317**, 513–516 (2007).
- Sun, B. *et al.* Impediment of *E. coli* UvrD by DNA-destabilizing force reveals a strained-inchworm mechanism of DNA unwinding. *EMBO J.* **27**, 3279–3287 (2008).
- Dessinges, M. N., Lionnet, T., Xi, X. G., Bensimon, D. & Croquette, V. Single-molecule assay reveals strand switching and enhanced processivity of UvrD. *Proc. Natl Acad. Sci. USA* **101**, 6439–6444 (2004).
- Chemla, Y. R. *et al.* Mechanism of force generation of a viral DNA packaging motor. *Cell* **122**, 683–692 (2005).
- Tsay, J. M., Sippy, J., Feiss, M. & Smith, D. E. The Q motif of a viral packaging motor governs its force generation and communicates ATP recognition to DNA interaction. *Proc. Natl Acad. Sci. USA* **106**, 14355–14360 (2009).
- Singleton, M. R., Sawaya, M. R., Ellenberger, T. & Wigley, D. B. Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. *Cell* **101**, 589–600 (2000).
- Patel, S. S., Rosenberg, A. H., Studier, F. W. & Johnson, K. A. Large scale purification and biochemical characterization of T7 primase/helicase proteins. Evidence for homodimer and heterodimer formation. *J. Biol. Chem.* **267**, 15013–15021 (1992).
- Mathews, C. K. Biochemistry of deoxyribonucleic acid-defective amber mutants of bacteriophage T4. 3. Nucleotide pools. *J. Biol. Chem.* **247**, 7430–7438 (1972).
- Hingorani, M. M. & Patel, S. S. Interactions of bacteriophage T7 DNA primase/helicase protein with single-stranded and double-stranded DNAs. *Biochemistry* **32**, 12478–12487 (1993).
- Liao, J. C., Jeong, Y. J., Kim, D. E., Patel, S. S. & Oster, G. Mechanochemistry of T7 DNA helicase. *J. Mol. Biol.* **350**, 452–475 (2005).
- Crampton, D. J., Mukherjee, S. & Richardson, C. C. DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. *Mol. Cell* **21**, 165–174 (2006).
- Yu, X., Hingorani, M. M., Patel, S. S. & Egelman, E. H. DNA is bound within the central hole to one or two of the six subunits of the T7 DNA helicase. *Nature Struct. Biol.* **3**, 740–743 (1996).
- Enemark, E. J. & Joshua-Tor, L. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**, 270–275 (2006).
- Thomsen, N. D. & Berger, J. M. Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* **139**, 523–534 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Wang laboratory for critical reading of the manuscript. We also thank M. A. Hall for assistance with single-molecule assays, data acquisition and data analysis. We wish to acknowledge support from National Institutes of Health grants (GM059849 to M.D.W.; GM55310 to S.S.P.), National Science Foundation grant (MCB-0820293 to M.D.W.) and Cornell's Molecular Biophysics Training Grant (T32GM008267) Traineeship (to D.S.J. and B.Y.S.).

Author Contributions B.S., D.S.J., S.S.P. and M.D.W. designed the experiments. D.S.J. found the helicase slippage with ATP. B.S. carried out all single-molecule work and, together with B.Y.S., analysed and interpreted single-molecule data. G.P. performed all the ensemble experiments. M.P. and G.P. purified and analysed the wild-type and mutant T7 gp4 proteins. M.D.W. formulated the theoretical models. B.S., D.S.J., B.Y.S., S.S.P. and M.D.W. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.D.W. (mwang@physics.cornell.edu) or S.S.P. (patells@umd.edu).

METHODS

Protein and DNA preparations. Wild-type T7 helicase (gp4A') and Y535F 4A' were expressed and purified as described previously¹⁷. A 5.2 kb DNA was constructed as described elsewhere^{9,25}, with minor modifications. Briefly, a ~1.1 kb anchoring segment was prepared by PCR from pRL574 using a digoxigenin-labelled primer, and then digested with BstXI (NEB) to produce a 3 bp overhang. A ~4.1 kb unzipping/translocation/unwinding segment was derived from pCP681 by digestion with EarI (NEB) and ligated to a biotin-labelled 37 bp segment lacking a 5' phosphate on the distal end. The anchoring segment and unzipping segment were then ligated, with a nick due to the missing phosphate. For ssDNA translocation experiments (Supplementary Fig. 8), the ~4.1 kb segment was capped with a hairpin (5'-TAGGGCGACCTAGCTCTATGCTAGGTCGCC-3').

Single-molecule assays. Sample preparation was similar to that previously described⁹. Briefly, helicase was prepared by first incubating 2 μ M of the helicase monomer for 20 min in the unwinding buffer. This solution was then further diluted to obtain the final experimental concentration of helicase monomer, nucleotides and MgCl₂. DNA tethers were formed by first non-specifically coating the sample chamber surface with anti-digoxigenin (Roche), followed by an incubation with digoxigenin tagged DNA. Streptavidin-coated 0.48 μ m polystyrene microspheres were then added to the chamber. Finally, helicase solution was flowed in just before data acquisition. The helicase unwinding buffer was 20 mM Tris-HCl (pH 7.5), 3 mM EDTA, 0.02% Tween 20, 50 mM NaCl, NTPs or dNTPs at the concentrations specified in the text, and MgCl₂ at a concentration 5 mM in excess of the total nucleotide concentration (Supplementary Fig. 13). The helicase monomer concentration was adjusted between 1–500 nM for each buffer condition so that the average unwinding initiation time (defined as the time between when the DNA was initially mechanically unzipped and when the helicase began to unwind) was approximately the same for all experiments (Supplementary Fig. 3).

Experiments were conducted in a climate-controlled room at a temperature of 23.3 °C, but owing to local laser trap heating the temperature increased slightly to 25 \pm 1 °C (ref. 26). Each experiment was conducted in the following steps (Supplementary Fig. 1). First, several hundred base pairs of dsDNA were mechanically unzipped, at a constant velocity of 1,400 bp s⁻¹, to produce a ssDNA loading region for helicase. Second, after the force dropped owing to helicase loading and initiation of unwinding, several hundred more base pairs were mechanically unzipped to generate ssDNA for helicase translocation. Third, the fork position was maintained until the force dropped again, indicating that the helicase had again reached the junction, at which point the force was allowed to drop to

8 pN and then maintained at this level as helicase unwound the remaining ~3 kb of dsDNA. Measurements of ssDNA translocation rates and dsDNA unwinding rates by T7 helicase were thus obtained for each tether.

Data collection and analysis. Data were low-pass filtered to 5 kHz and digitized at 12 kHz, then were further averaged to 110 Hz. The acquired data signals were converted into unwound base pairs as previously described^{9,25}. To improve positional accuracy and precision, the data were then aligned to a theoretical unzipping curve for the mechanically unzipped section of the DNA²⁷. Slippage events were identified by a threshold on the instantaneous unwinding rate at each sequence position (Supplementary Fig. 4). We used a threshold of 2,000 bp s⁻¹ in the reverse velocity for identifying slippage. Unwinding rates from each trace were found from linear fits to the unwinding between adjacent slippage events. An average unwinding rate was obtained from a number of traces. Distances travelled between slips were compiled to determine processivity. These distances followed an exponential distribution, indicating a stochastic process in slippage²⁸. Processivity is defined as the mean distance of the distribution (Supplementary Fig. 4b).

Modeling. If dTTP and ATP compete for binding to helicase according to the kinetic pathway outlined in Fig. 2d, then the resulting unwinding rate is:

$$V_{\text{tot}} = \left(V_{\text{max}}^{\text{ATP}} \frac{[\text{ATP}]}{K_M^{\text{ATP}}} + V_{\text{max}}^{\text{dTTP}} \frac{[\text{dTTP}]}{K_M^{\text{dTTP}}} \right) / \left(1 + \frac{[\text{ATP}]}{K_M^{\text{ATP}}} + \frac{[\text{dTTP}]}{K_M^{\text{dTTP}}} \right),$$

where for each type of nucleotide $K_M = \frac{k_{-1} + k_2}{k_1}$ and $V_{\text{max}} = sk_2$ with s being the step size (in nucleotides) (see Supplementary Discussion). In the presence of dTTP and ATP, if n helicase subunits coordinate in their chemo-mechanical activities and DNA binding, then the resulting distance between slips (processivity) is:

$$d_{\text{processivity}} = c \left(V_{\text{max}}^{\text{ATP}} \frac{[\text{ATP}]}{K_M^{\text{ATP}}} + V_{\text{max}}^{\text{dTTP}} \frac{[\text{dTTP}]}{K_M^{\text{dTTP}}} \right) / \left(\frac{[\text{ATP}]/K_M^{\text{ATP}}}{[\text{ATP}]/K_M^{\text{ATP}} + [\text{dTTP}]/K_M^{\text{dTTP}}} \right)^{n-1}$$

(Supplementary Discussion), with c being a proportionality constant. This expression was used to fit data in Fig. 3b, c with c and n as fit parameters.

25. Koch, S. J., Shundrovsky, A., Jantzen, B. C. & Wang, M. D. Probing protein-DNA interactions by unzipping a single DNA double helix. *Biophys. J.* **83**, 1098–1105 (2002).
26. Peterman, E. J., Gittes, F. & Schmidt, C. F. Laser-induced heating in optical traps. *Biophys. J.* **84**, 1308–1316 (2003).
27. Shundrovsky, A., Smith, C. L., Lis, J. T., Peterson, C. L. & Wang, M. D. Probing SWI/SNF remodeling of the nucleosome by unzipping single DNA molecules. *Nature Struct. Mol. Biol.* **13**, 549–554 (2006).
28. Lohman, T. M., Tornko, E. J. & Wu, C. G. Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nature Rev. Mol. Cell Biol.* **9**, 391–401 (2008).

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature10459

Detection of prokaryotic mRNA signifies microbial viability and promotes immunity

Leif E. Sander, Michael J. Davis, Mark V. Boekschoten,
Derk Amsen, Christopher C. Dascher, Bernard Ryffel,
Joel A. Swanson, Michael Müller & J. Magarian Blander

Nature **474**, 385–389 (2011).

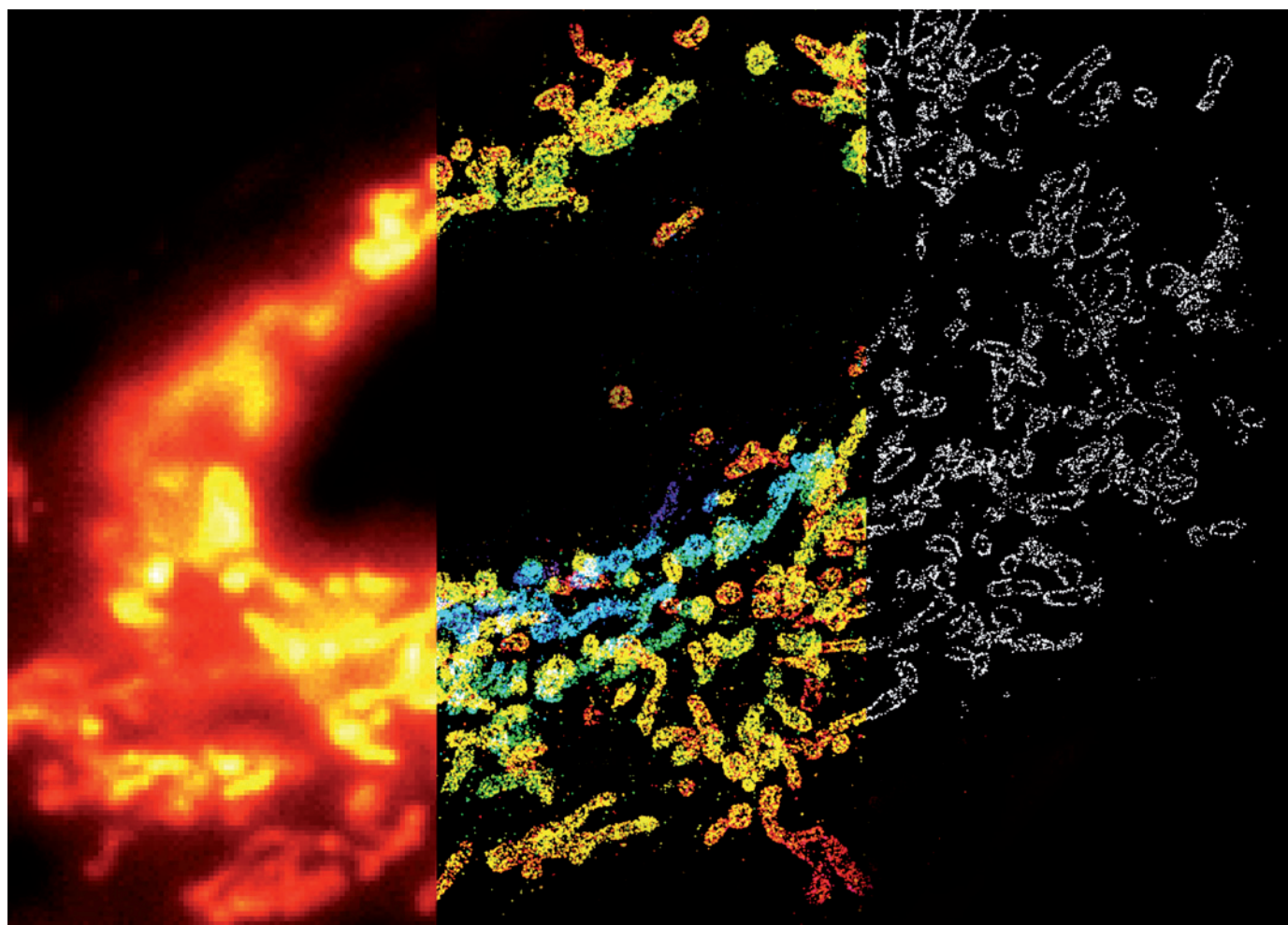
In Fig. 1d of this Letter, the labels HKEC and EC were swapped in the print version. The lane labelled HKEC should be labelled EC and the lane labelled EC should be labelled HKEC. The error has been corrected online in the HTML and PDF versions.

TECHNOLOGY FEATURE

BRIGHT LIGHT, BETTER LABELS

The tiniest structures in cells can be seen only using sophisticated instrumentation and informatics, but what biologists really need are improved fluorescent probes.

ZHUANG LAB/HARVARD UNIV.



Mitochondria in a cell, imaged by conventional microscopy (left), and super-resolution microscopy colour-coded by depth (middle) and in cross-section (right).

BY MONYA BAKER

Scientists love to decorate their favourite biomolecules with fluorescent tags. Attaching light-emitting labels to a protein can reveal when and where in a cell it functions, but usually the details are fuzzy. Optical microscopes use light with wavelengths between 350 and 750 nanometres, and structures smaller than about 200 nm cannot be seen clearly. That is much bigger than the thickness of a cell membrane and is about half as long as the mitochondria that supply cells' energy. At this scale, many cellular secrets are invisible. The protein machinery

that allows a virus to invade a cell is blurry, as are the synapses across which neurons communicate.

The past few years have seen the rise of a suite of techniques, collectively known as super-resolution microscopy, that can use light to reveal structures much smaller than the theoretical limit. The trick is to control fluorescent labels, or fluorophores, so that not all of them signal at once. Light from each individual fluorophore creates a blur, but as long as blurs don't overlap, they can be resolved into individual points at their centres. This allows the position of the fluorophore to be identified precisely, revealing features as small as 20 nm.

"The super-resolution that we have developed doesn't rely on changing the wave nature of light," says Stefan Hell, director of nanobiophotonics at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany. "It relies on turning dyes on and off."

Although advances in instrumentation and informatics should not be overlooked, many researchers believe that it is better-performing fluorescent labels that will allow super-resolution microscopy to continue to move forward. "That's an area where the field will see the biggest advances," says Jan Liphardt, a biophysicist at the University of California, Berkeley. "That's been limiting all of us."

Electron microscopes can resolve features less than a nanometre long — even smaller than super-resolution. But electron microscopy requires elaborate preparation of samples: usually, cells must be ‘fixed’ with preservatives and then embedded in resin or frozen. By contrast, many forms of super-resolution microscopy can be done with live cells. And with fixed cells, labels for optical microscopy can identify proteins more specifically than can those available for electron microscopy.

Most super-resolution techniques fall into two categories. In one, sometimes called illumination-based super-resolution, precise geometric patterns of light shine repeatedly across a sample to control which fluorophores are active. In the other, sometimes called probe-based super-resolution, conditions are tuned so that just a few fluorophores emit light at a time.

Whereas illumination-based super-resolution microscopy requires specialized optical equipment, probe-based techniques do not. Experiments using the latter technique are relatively easy to set up (see ‘Starting up in super-resolution’). However, only a few



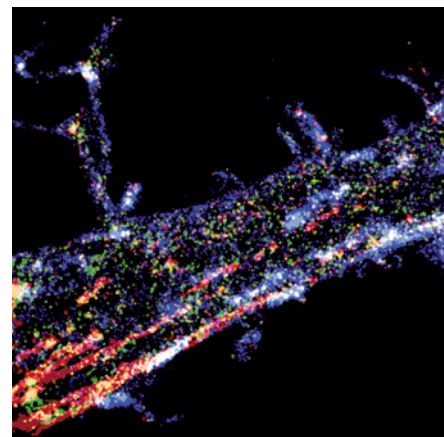
“You make a map of where the molecules were. That’s your image.”

Sam Hess

dozen of the hundreds of extant fluorescent proteins and dyes have the requisite properties for probe-based super-resolution microscopy: the ability to change from one ‘spectral state’ to another when exposed to certain wavelengths of light (see ‘Fluorescent proteins for super-resolution microscopy’). Some remain dark until they are activated; others go from one colour to another.

ACRONYM UPROAR

Three labs independently developed the first probe-based techniques in 2006: fluorescence photoactivation localization microscopy (fPALM) was described¹ by Sam Hess, a physicist at the University of Maine in Orono; photoactivated localization microscopy (PALM) was described² by Eric Betzig and Harald Hess, physicists at the Howard Hughes Medical Institute’s Janelia Farm Research Campus in Ashburn, Virginia; and stochastic optical reconstruction microscopy (STORM) was described³ by Xiaowei Zhuang, a physicist at Harvard University in Cambridge, Massachusetts. Perhaps one of the most confusing aspects of these and other probe-based techniques is what to call them. Commonly used terms include fPALM/STORM and variations such as single-molecule localization microscopy (SMLM) and single-molecule active-control microscopy (SMACM). But the underlying concepts behind all the probe-based techniques are the same, says Sam Hess. “You somehow control the molecules so you only have a few visible at a time, you find their



Live-cell super-resolution images showing how actin and membrane proteins associate.

position, you cycle through a whole bunch of molecules, and you make a map of where the molecules were. That’s your image.”

These techniques require more control over fluorophores than most scientists are used to, says Michael Davidson, director of the optical microscopy division at the National High Magnetic Laboratory in Tallahassee, Florida. “A lot of people are jumping into this. I get the question of what probes to use probably ten times a week,” he adds.

Fluorescent proteins are commonly used for super-resolution microscopy. The genes that code for them, often taken from jellyfish or other sea creatures, are fused with the genes for the proteins being studied, so that when

Starting up in super-resolution

Interest in super-resolution techniques is widespread, but relatively few labs have taken the plunge. Here are some tips.

Optimize conditions first. “Before we go through collaboration, we ask that people try these labels out first under a regular fluorescent microscope,” says Harald Hess, a physicist at the Howard Hughes Medical Institute’s Janelia Farm Research Campus in Ashburn, Virginia. A fluorescent protein that works seamlessly with one protein of interest could completely disrupt another. “Make sure the proteins photoactivate, make sure the cell health is okay, make sure that the density’s right,” says Hess. When a super-resolution experiment doesn’t work as expected, researchers can be quick to blame the optical equipment. Often the problem is actually with the biological label.

Watch out for artefacts that are no longer invisible. A 50-nanometre perturbation is invisible in conventional microscopy experiments. But in super-resolution, that

distance can tell you whether two proteins cluster together or stay apart. Artefacts that researchers could once safely ignore — microscope drift, a label’s slight effects on localization — must now be considered.

Get to know your fluorophore. Researchers can’t predict from the literature how a protein will behave in their hands. Even when fused to the same protein, a fluorophore’s photostability — the number of photons it can give off — can vary from instrument to instrument, sample to sample and culture to culture, especially under differing oxygen levels. “The numbers will be really specific to your experimental conditions,” says Robert Campbell, a protein engineer at the University of Alberta in Edmonton. The order of relative photostability of fluorophores should remain the same, says Campbell, “but I wouldn’t bet my life on it”.

Consider your lasers. Xiaowei Zhuang, a physicist at Harvard University in Cambridge, Massachusetts, developed stochastic optical

reconstruction microscopy (STORM) using a low-powered laser to avoid damaging the sample, but it took so long to switch the fluorophores that acquiring an image took several minutes. With more powerful lasers, transitions occur in a millisecond. Michael Davidson, director of optical microscopy at the National High Magnetic Laboratory in Tallahassee, Florida, recommends at least 100 milliwatts for green wavelengths, and up to 200 milliwatts for far-red. The Laser Combiner produced by Agilent of Santa Clara, California, contains four lasers, putting several wavelengths under easy control.

Don’t overactivate the probes. Setting up a probe-based super-resolution experiment is easy, but calculating localization spots does not guarantee resolution higher than that of conventional microscopy. If too many probes are activated, localization represents not individual molecules, but an average of several, explains Zhuang. “Just because you get a STORM-like type of image doesn’t mean that it has high quality,” she says. **M.B.**

the proteins are produced, they too are joined. Thus, when the fluorophore probe lights up, it allows researchers to locate the studied protein. The most popular protein for probe-based super-resolution microscopy is probably mEos2 (ref. 4). When first expressed, it fluoresces green, but a burst of ultraviolet light turns it red. Such 'photoconvertible' fluorophores offer certain advantages over those that start out in a dark, non-fluorescent state: they allow researchers to image the protein before experiments begin, and so more easily pick out healthy cells that are producing high levels of the labelled protein. What is more, newly produced proteins are different colours from those that have already been imaged, so researchers can follow pools of proteins over time and get a sense of their rates of production and destruction (see 'How to build a fluorescent protein').

EVERYTHING IS ILLUMINATED

Often, though, one fluorophore per experiment is not enough. "Most of the outstanding questions that people want nanometric accuracy for are in the relationship of two or more different proteins relative to each other," explains Jennifer Lippincott-Schwartz, a cell biologist at the National Institutes of Health in Bethesda, Maryland, and part of the team that invented PALM. "The only way that you can address that is using different markers at the same time."

Gleb Shtengel, a physicist at Janelia Farm, says that getting two labels to work together inside a cell is difficult, partly because the optimal conditions for each are not always the same. Fluorophores always prove trickier to work with than the imaging apparatus. "You have to add another laser, but that's the simplest part," says Shtengel. Putting the brighter label on the less-expressed protein can help to make sure that enough data can be collected on each of the proteins of interest to fix their locations definitively; expression levels must also be sufficient and reliable for both proteins.

And then there are the spectral considerations. If researchers want to use a second label alongside mEos2, for example, they have to find one unaffected by both red and green wavelengths of light. A protein described⁵ this year

could be a big help: it converts from orange to far-red, a much desired colour that is distinct from both the natural fluorescence of cells and that of other popular fluorophores. "The palette is so small right now that any addition is a big step forward, especially if you add a colour in part of the spectrum that's empty," says Shtengel.

But researchers are succeeding in using two-colour super-resolution microscopy. That has allowed them to address questions such as whether cell-surface receptors implicated in cancer are randomly assorted or are co-localized on the plasma membrane. Lippincott-Schwartz has described^{6,7} a general technique that allows researchers to quantify how proteins cluster together on plasma membranes, and to assess the size, abundance and density of clusters.

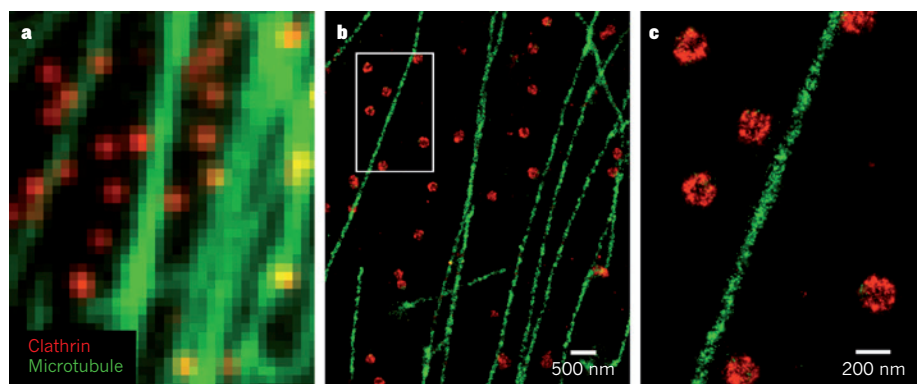
Even illumination-based techniques are benefiting from new fluorophores. A technique called stimulated emission depletion works by pairing lasers: one excites a spot to fluoresce,

"A lot of people are jumping into super resolution. I get the question of what probes to use probably ten times a week."

and the other shrinks the area of fluorescence by further exciting fluorophores on its periphery into a special dark state. To collect an image, the paired laser beams scan across the sample, repeatedly applying intense beams of

light that force fluorophores into the appropriate state but can also damage cells. Hell and his colleagues last month described⁸ a fluorescent protein that can enable illumination-based super-resolution microscopy in extremely low light levels. Although most fluorescent proteins bleach out, or lose their fluorescence, with repeated imaging, this new protein can be switched on and off more than 1,000 times. The researchers were able to image dendritic spines (signal-receiving outgrowths on neurons) at light levels one million times lower than had previously been documented, and the technique can work with a standard confocal microscope, says Hell.

Lippincott-Schwartz and others are working out ways to make conventional fluorophores



Conventional (a) and super-resolution (b, c) microscopy of microtubules and clathrin protein clusters.

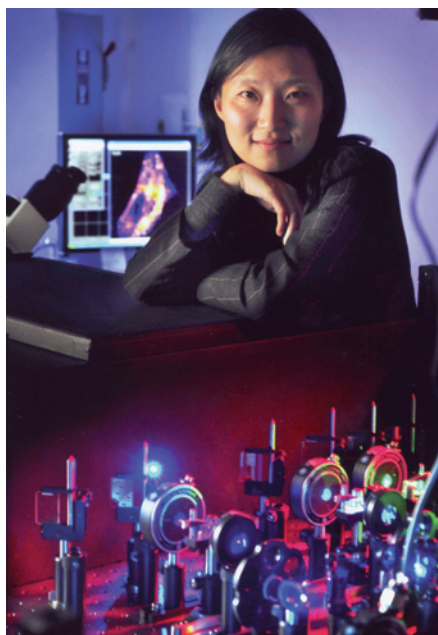
ZHUANG LAB/HARVARD UNIV.

amenable to probe-based super-resolution microscopy. Instead of lighting up just a few molecules at once, they activate an entire population and wait for the fluorescent proteins to slowly turn off. The analysis identifies the loss of signal, she explains. “As they bleach, molecules switch off and leave a hole that can be fit to determine where the molecule was.” Data for localizing dark holes are much noisier than those for localizing bright points, but the technique allows researchers to work with several labels at once. In unpublished work, Lippincott-Schwartz has been able to visualize as many as four fluorophores in a single fixed sample, and she thinks that the technique can also be made to work in live cells.

DESIRABLE DYES

To many cell biologists, the term fluorescent label is synonymous with fluorescent protein, but there are also small-molecule fluorescent dyes. Dyes tend to be more photostable than fluorescent proteins, so they can emit an order of magnitude more photons, which means that dye molecules can be detected and pinpointed more reliably, explains Markus Sauer, a biophysicist at the University of Würzburg in Germany. “The higher photon yield goes in hand with higher localization precision and thus a higher optical resolution,” he says.

The speed at which dyes turn on and off is also an advantage. In the first demonstration of live-cell, three-dimensional STORM, Zhuang used six different probes: four dyes and two proteins⁹. One of the dyes, Alexa 647, allowed an image to be taken in one second; proteins required substantially longer, at 30 seconds per image. Collecting more images in less time is a practical advantage for all samples, particularly



Xiaowei Zhuang of Harvard University looks for tiny details using fluorescent proteins and dyes.

for live cells, says Zhuang. “If you can’t switch the probes fast, you can only image slow processes,” she adds.

The problem is that dyes are often less convenient than fluorescent proteins. Whereas researchers can label proteins with fluorescent proteins by introducing genes into cultured cells, dyes have to be attached in a separate step. The most common technique is to combine them with antibodies against a protein of interest. Usually, researchers label ‘secondary antibodies’, which themselves attach to antibodies against the protein of interest — a practice that

allows the same reagents to be used in multiple experiments. However, because it is the antibody rather than the protein that is visualized, the dyes are somewhat distant from the protein of interest.

Antibodies can usually be used only on fixed cells, but they do have advantages. Relevant techniques are in common use, and antibodies work in samples that can’t be transfected, such as human biopsies. What is more, the target proteins are produced naturally, rather than from introduced genes, which can have aberrant expression. Last year, Zhuang and her colleagues reported¹⁰ that they had used labelled antibodies with STORM to interrogate the locations of ten different proteins within synapses, distinguishing which occurred on the signal-sending (pre-synaptic) side and signal-receiving (post-synaptic) side — something that would be impossible in conventional microscopy because the synapse is so small.

There are also ways to use dyes without antibodies: ‘soluble ligands’, or secreted proteins that attach to cell surfaces can be produced, labelled and then added to cell cultures directly. Intracellular proteins can be labelled using a ‘hybrid-fusion’ approach. Instead of being fused to a fluorescent protein, a protein of interest is joined to a ‘protein hook’ that can attach to the dye molecules. A variety of tags are in use, and the technique can even work with commercially available chemical-tag kits made for conventional microscopy¹¹. But the dye can sometimes attach to biomolecules other than the target, says Robert Campbell, a protein engineer at the University of Alberta in Edmonton. “That raises up background fluorescence, and that limits the level at which you can see the protein.”

How to build a fluorescent protein

To be useful for super-resolution microscopy, a fluorescent protein must have all the properties necessary for standard imaging: it can’t be toxic; it must label the intended target; it must be biologically active at the same temperatures as mammalian cells (no small feat, given that these proteins generally come from sea creatures living in chilly waters); it must be bright; and its fluorescence must stand out from background.

Such proteins are often taken from jellyfish, anemones or coral, but they all have the same general shape: a flat sheet rolled up into a ‘ β -barrel’, with a helix spiralling into the centre. Three amino acids at the end of the helix create the chromophore — the part responsible for the fluorescence. For a protein to be photoactivatable, the parts in or around the chromophore must undergo reactions catalysed by light. Swapping in different amino acids can create new colours

and set the stage for light-activated reactions that change the properties of the protein.

Even if a protein has the desired colour and photoactivity, it may not be bright or well-behaved enough to be useful in microscopy, so researchers use random mutagenesis to hunt for beneficial mutations over the entire barrel. Super-resolution imaging makes heavy demands on proteins, says Vladislav Verkhusha, a structural biologist at Albert Einstein College of Medicine in New York City, who has made many photoactivatable proteins, including the first far-red one⁵. “If you want to have tenfold better resolution, you need 100-fold photostability,” he says.

And for reasons that aren’t entirely clear, the way that proteins behave in a group does not perfectly represent their behaviour at the single-molecule level, explains George Patterson, a physicist at the US National Institutes of Health in Bethesda, Maryland,

who helped to develop the first practical photoactivatable fluorescent protein and a super-resolution technique².

Fluorescent proteins occur naturally as bulky tetramers, impractical for labels, so research groups have to modify them. One group might break the four-barrel proteins into individual stable barrels more suitable for labelling; another might shift the protein’s colour spectrum; another might make it photoactivatable; and yet others might make more general improvements.

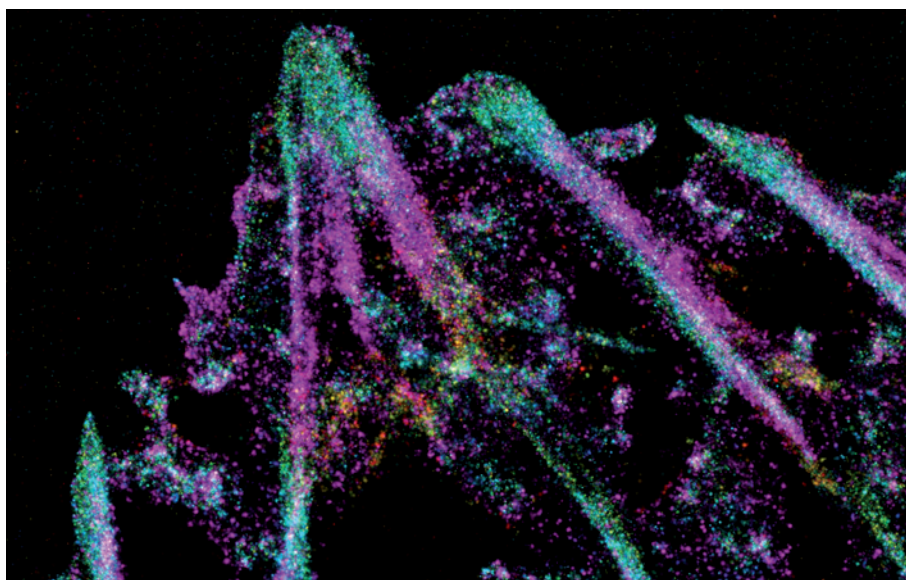
Researchers just can’t get enough. Now more than ever, the better the fluorophore, the more biology it can reveal. “The fluorophore is at centre stage of the whole development,” says Stefan Hell, a nanomicroscopist at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany. “The fluorophore is decisive. It allows you to get the pictures.” **M.B.**

Improved analysis will also help scientists to get more from their labels. Researchers led by Sam Hess showed¹² that three fluorophores that all emit in the orange-red wavelengths could be distinguished from each other. Conventional microscopy would not be able to separate, say, a greenish-yellow label that emits two green photons for each red one from an orangish-yellow label that emits one green photon for each red one, but super-resolution microscopy can distinguish such signals because emitted photons are attributed to individual proteins. In such a case, “it’s okay to have the emission spectrum overlap because you are imaging individual molecules”, says Hess. His team was able to use three labels with overlapping spectra to simultaneously image two membrane proteins and a cytoskeleton protein, showing how these different components of the cell interact.

BETTER RESOLUTION THROUGH COMPUTATION

Better analysis and more sophisticated algorithms should also help researchers who are using only one label at a time. To speed imaging, researchers would like to increase the number of fluorophores that emit light at any given time. But if too many fluorophores emit too close together, their signals overlap and cannot be resolved into individual points. Several groups are working on software that lets scientists image more labels in smaller spaces. For example, researchers at the University of Oxford, UK, adapted¹³ an algorithm originally developed to study crowded star systems, and used it in probe-based super-resolution microscopy. They showed that it could detect more fluorophores than could two imaging algorithms commonly used in microscopy.

Aleksandra Radenovic, a biophysicist at the Swiss Federal Institute of Technology in Lausanne, has designed computational approaches to mitigate artefacts caused when ‘bleached’ proteins, which have supposedly lost their fluorescence permanently, revert to a state in which they can be activated¹⁴. The effort grew out of another project, exploring dense protein clusters on the cell membrane. After a protein fragment chosen as a negative control displayed unexpectedly high levels of clustering,



H. HESS/C. WATERMAN/M. DAVIDSON

Super-resolution imaging reveals the molecular architecture that enables cellular adhesion.

Radenovic and her co-workers studied the activation times of individual molecules of mEos2. The data showed that signals from similar locations clustered together in time. The sequence of signalling molecules should be random across a sample, so these results indicated that the same protein was signalling more than once and was being misinterpreted as multiple proteins, explains Radenovic. “Just looking at the time domain, you can get rid of those artefacts,” she says.

Although probe-based super-resolution microscopy can be done using standard fluorescence microscopes, several manufacturers offer systems built specifically for this purpose, along with software for analysing the data. Such microscopes are designed to optimize the activation of probes. Licensing agreements restrict which acronyms each manufacturer uses in marketing, but a machine that works for one form of probe-based microscopy generally works for other forms as well. Tokyo-based company Nikon has installed its system in dozens of labs; Leica Microsystems of Wetzlar and Zeiss of Oberkochen, both in Germany, have also introduced systems. And small start-up companies, such as Vutara in

Salt Lake City, Utah, are getting into the market as well. Applied Precision of Issaquah, Washington (acquired in April by GE Healthcare of Fairfield, Connecticut), plans to roll out its probe-based super-resolution system, Monet, later this year. Most of these companies also make instruments for illumination-based microscopy, which require specialized components.

With or without dedicated instruments, researchers are keen to try their hand at super-resolution microscopy. So far, most papers demonstrate proof of principle for microscope methods rather than fundamental new biology uncovered by the techniques, but the balance is shifting, says Davidson. “It’s going to be an explosive field. It’s just now raising its head, and it’s about to take off like a bat out of hell.” ■

Monya Baker is technology editor for *Nature* and *Nature Methods*.

- Hess, S. T., Girirajan, T. P. K. & Mason, M. D. *Biophys. J.* **91**, 4258–4272 (2006).
- Betzig, E. *et al. Science* **313**, 1642–1645 (2006).
- Rust, M. J., Bates, M. & Zhuang, X. *Nature Meth.* **3**, 793–796 (2006).
- McKinney, S. A., Murphy, C. S., Hazelwood, K. L., Davidson, M. W. & Looger, L. L. *Nature Meth.* **6**, 131–133 (2009).
- Subach, O. M. *et al. Nature Meth.* **8**, 771–777 (2011).
- Subach, F. V., Patterson, G. H., Renz, M., Lippincott-Schwartz, J. & Verkhusha, V. V. *J. Am. Chem. Soc.* **132**, 6481–6491 (2010).
- Sengupta, P. *et al. Nature Meth.* <http://dx.doi.org/10.1038/nmeth.1704> (2011).
- Grotjohann, T. *et al. Nature* <http://dx.doi.org/10.1038/nature10497> (2011).
- Jones, S. A., Shim, S.-H., He, J. & Zhuang, X. *Nature Meth.* **8**, 499–505 (2011).
- Dani, A., Huang, B., Bergan, J., Dulac, C. & Zhuang, X. *Neuron* **68**, 843–856 (2010).
- Klein, T. *et al. Nature Meth.* **8**, 7–9 (2011).
- Gunewardene, M. S. *et al. Biophys. J.* **101**, 1522–1528 (2011).
- Holden, S. J., Uphoff, S. & Kapanidis, A. N. *Nature Meth.* **8**, 279–280 (2011).
- Annibale, P., Vanni, S., Scarselli, M., Rothlisberger, U. & Radenovic, A. *Nature Meth.* **8**, 527–528 (2011).

FLUORESCENT PROTEINS FOR SUPER-RESOLUTION MICROSCOPY

Protein	Spectral states	
mEos2; tandem dimer Eos	Green to red	Brightest photoactivatable proteins described so far. mEos2 behaves well when fused with target proteins. (Dendra2 and KikGR also convert from green to red, but not as well)
PS-CFP2	Cyan to green	Can be used alongside red and orange proteins
PSmOrange	Orange to far-red	First photoconvertible protein in the advantageous far-red part of spectrum
PAmCherry	Dark to orange	Along with PAtag RFP, can be used in combination with dark-to-green proteins
PA-GFP	Dark to green	Behaves well when fused with target proteins. First photoactivatable protein, but not as bright as others
Dronpa	Dark to green to dark (reversible)	The ability to switch from dark to light and back to dark allows tracking and live-cell applications

CAREERS

COLUMN Every scientist needs compassion and support in the lab **p.145**

CAREERS BLOG The latest discussions and news on research jobs go.nature.com/z8g4a7

NATUREJOBS For the latest career listings and advice www.naturejobs.com



J. ENDICOTT/IMAGES.COM/CORBIS

EDUCATION

Inspiration for informatics

Trainees in bioinformatics and computational biology should seek depth of knowledge over breadth.

BY VIRGINIA GEWIN

This January, Alexander Sczyrba and his colleagues published what was at the time the largest metagenome ever assembled (M. Hess *et al. Science* **331**, 463–467; 2011). Collecting and collating genetic material from environmental samples is always a challenge; in this case, the metagenome came from parts of a cow's stomach, and contained more than 27,000 biomass-degrading genes and 15 microbe genomes. It totalled 268 gigabases. "We had to develop new algorithms to run analyses on computer clusters, or clouds, as using traditional methods would have taken 80 years on a single computer," says Sczyrba.

Sczyrba wants to focus his career on similar complex, leading-edge analyses. But the path hasn't been straightforward; when he was looking for a postdoc in 2008, it was tough to find institutions that could generate or analyse such large data sets. He landed a post at the US Department of Energy Joint Genome Institute (JGI) in Walnut Creek, California: a large-scale sequencing facility that offered access to data, computing resources and brain power. In 2010 alone, the JGI sequenced 170 metagenomes.

Soon, however, big sequencing centres won't be the only sources of data. "With next-generation sequencing, everybody can produce sequences; it's the analysis that is getting more important," says Sczyrba. Modern biologists need to be able to manage large data sets and explore new computational tools.

FINDING A PATH

Qualified candidates are hard to find, say recruiters in both industry and academia. That may be because, so far, there hasn't been a typical career path for bioinformaticians or computational biologists. "Often we find that it's the people motivated to simply roll up their sleeves and figure out on their own how to work with these data that have the strongest skills," says Jim Bristow, deputy director of programmes at the JGI. As more departments are established, the often circuitous routes once required to attain such skills will probably be replaced by more direct paths. The challenge is finding a training programme that will help researchers to keep pace in a rapidly changing, technology-driven field.

By conventional definitions, bioinformaticians develop new ways to acquire, organize and analyse biological data, whereas computational biologists develop mathematical models or simulation techniques to work out the ►

► data's biological significance. But these lines are blurring, and departments and training programmes are both proliferating and combining the fields.

"The demand for computational-biology training that we have today is way more than was expected a decade ago," says Burkhard Rost, president of the International Society for Computational Biology, which is based in La Jolla, California.

NOT JUST SKIN DEEP

The most obvious training route — pursuing an undergraduate degree in bioinformatics — isn't necessarily the best for a budding researcher. Some undergraduate programmes fail to provide the depth of knowledge sought by employers. "Often these trainees come with great-looking CVs, but when we press them on what they are capable of doing, they tend to be rather weak," says Nick Goldman, research and training coordinator at the European Bioinformatics Institute in Hinxton, UK. Goldman is most impressed by applicants who have actively pursued training in both informatics and the area of research in which they're interested — for example, someone with a computing degree who has done a molecular-biology project (see "Talent checklist").

Goldman says that students should be wary of learning about only the latest software or genome-mining tool, without gaining a full understanding of the biological topics. Recruiters want savvy scientists who understand technology's ability to address questions. Steve Cleaver, head of quantitative biology at Novartis Institutes for BioMedical Research in Cambridge, Massachusetts, says that the key to a sustainable career in the field is the ability to turn a scientific question into a statistical hypothesis. "But those who can ride the tech waves are well positioned to find career success," he adds. Without a doubt, he adds, the next generation of biologists will be more conversant in bioinformatics. "It's all about cross-training — getting the appropriate training in both analytical science and biology during graduate school to make a meaningful contribution," says Cleaver.

Picking a programme with comprehensive training modules in statistics, computer science and/or biology can be an effective strategy. But Søren Brunak, director of the Center for Biological Sequence Analysis at the Technical University of Denmark in Lyngby, says that researchers should avoid training programmes that focus on just a few data types. With the expansion in high-throughput sequencing of genomes, proteins and metabolites, programmes that focus on a single area, such as genomics, don't adequately prepare students for the job market, says Brunak. "Analyses conducted now are much more reliant on combinations of data types — for example, combining molecular-level data with patient records — than they were before," he notes.

Aspiring principal investigators can go one step further to find the best graduate training for the career they want, by deciding whether to focus on developing tools, such as algorithms to analyse data, or applying those tools to turn data into knowledge.



"We don't know where we'll be in ten years because the technologies and ideas are moving so fast."

Alexander Sczyrba

to shape the tool developers. It accepts only candidates who demonstrate a core strength in an analytical field such as computer science or maths, or have a dual degree combining one of these fields with biology. Christopher Lee, director of the programme, says that many bioinformatics courses are affiliated with data-rich biology labs on campus, supplying the students needed to tackle a flood of data. They often lack, however, the matrix of expertise necessary to conduct innovative analyses. Lee hopes that the UCLA programme will foster such expertise.

A few graduate training programmes, notably those at the Netherlands Bioinformatics

Center in Nijmegen, cater to students with backgrounds in either computer science or biology. "We want to train the tool shapers as well as the people more into applying the tools in a biological setting," says Celia van Gelder, the centre's education project leader. "Over the past 10–20 years, the field of biology has become more computational, with bioinformatics serving as an interdisciplinary field that links researchers who can't otherwise readily talk to one another." The scope of work is widening, she says. As a result, demand for bioinformatics training continues to increase across Europe — with greater emphasis placed on data analysis at all levels. "We produce trainees who have multidisciplinary training in molecular-biology principles as well as algorithms to deal with data," says Jaap Heringa, the centre's scientific director for bioinformatics education. "Things move so fast in bioinformatics, we are constantly innovating our courses," he adds. Murphy agrees; Carnegie Mellon and the University of Pittsburgh offer in-depth training. "We are pretty clear in the application materials that our programme is not for people who want to get enough of a smattering of computational biology to get a job," says Murphy.

The University of California, Los Angeles (UCLA), has a bioinformatics PhD programme designed

to shape the tool developers. It accepts only candidates who demonstrate a core strength in an analytical field such as computer science or maths, or have a dual degree combining one of these fields with biology. Christopher Lee, director of the programme, says that many bioinformatics courses are affiliated with data-rich biology labs on campus, supplying the students needed to tackle a flood of data. They often lack, however, the matrix of expertise necessary to conduct innovative analyses. Lee hopes that the UCLA programme will foster such expertise.

EXPANDING OPTIONS

This trend towards creating more comprehensive, interdisciplinary training programmes has gained momentum at biology strongholds in the United States. In July 2010, Dartmouth Medical School in Hanover, New Hampshire, established the Institute for Quantitative Biological Sciences in nearby Lebanon. Its graduate offerings combine modules in bioinformatics, biostatistics and epidemiology. "We have created what we think is a model of the future — training computational-biology students to speak multiple languages beyond bioinformatics," says the centre's director, Jason Moore. He adds that the key is assuming complexity rather than simplicity when approaching a problem.

In August, Moore secured funding to create a US National Institutes of Health (NIH) Center for Biomedical Research Excellence, through which he will mentor five early-career bioinformatics faculty members, to be recruited over the next 3–4 years. After two years of learning how to secure competitive funding, among other things, trainees will be required to submit an application for an R01 grant, the NIH's main funding mechanism. "We really want to provide a well rounded education so that our new recruits can secure funding for — and conduct — well designed studies in computational biology," says Moore.

Other medical schools are also taking the plunge. Duke University School of Medicine in Durham, North Carolina, formed its Department of Biostatistics and Bioinformatics in 2000. This year, it opens its first master's programme, says Elizabeth Delong, chair of the department.

BASIC SKILLS

Talent checklist

- Be at least conversant in the broad range of disciplines contributing to bioinformatics — from statistics to molecular biology to computer science.
- Most work, especially in industry, is done in teams, so communication skills are always in demand.
- Get experience in handling massive data sets. Learn to parse data or run analyses in parallel — using, for example, cloud computing.
- Learn to write programmes in software languages such as Perl or R.
- Cultivate a deep knowledge of at least one area of biology. **V.G.**

And in September, the University of Michigan Medical School in Ann Arbor established a computational-medicine and bioinformatics department to help attract new faculty members and trainees. In June, Emory University School of Medicine in Atlanta, Georgia, launched a biomedical-informatics department with the goal of combining expertise in imaging, computer science and biology to improve patient care. It will recruit four or five researchers over the next few years. "Our particular strength is training computer scientists who want to transition into biomedical informatics, and bringing them together with clinicians to use informatics to treat disease," says department chair Joel Saltz.

Qualified postdocs remain in demand. "It can be very difficult for individual investigators to hire a postdoc in bioinformatics," says Tom Tullius, interim chair of the bioinformatics programme at Boston University in Massachusetts. He attributes the paucity of candidates in part to efforts over the past several years to build large teams at high-powered institutes — such as the Broad Institute in Cambridge, Massachusetts, or the Wellcome Trust Sanger Institute in Cambridge, UK

— leaving smaller labs struggling to find talent. The growth of training programmes could ease this.

Now sequencing centres won't be the sole providers of data, individual researchers, particularly at medical centres, will have ample data to fuel research and training. "We've passed out of the period of genome projects where

there were amazing public data raining down from the heavens; it's now possible to do exciting work without being associated with data-generating centres," says Lee.

Sczyrba, who begins a junior faculty position in metagenomics at the University of Bielefeld Center for Biotechnology in Germany this autumn, says that unpredictability is what makes the discipline so exciting. "We don't know where we will be in ten years because the technologies and ideas are moving so fast," he says. As Cleaver notes: "Perhaps the best career strategy is to stay flexible and curious." ■

Virginia Gewin is a freelance writer in Portland, Oregon.



BEAU LARK/CORBIS

COLUMN

The human touch

A little empathy goes a long way in the competitive confines of a laboratory, argues **Lydia Soraya Murray**.

As almost every scientist knows, a person's first year in research is an emotional minefield. One minute you're flying high. The next, you're banging your head against the wall, resisting the urge to draw in results with a marker pen. Forget the F-word; in science, it's the O-word that generates dread. I sometimes think that 'optimizing' should be spelled 'r.e.p.e.a.t.e.d.f.a.i.l.u.r.e.s'.

After a stimulating yet often soul-destroying start to my PhD, I have decided that coming to terms with the lows is one of the most important things that you can take away from your first year. Never mind the dreaded literature review; this is unquestionably more important.

Because scientists do incredibly specialized and often misunderstood work, it can be hard for people outside our particular fields to empathize with our attachment to our projects. I have a close friend who is a physician. After several weeks of my hard work culminated in what can be described only as 'diddly squat', my friend offered these consoling words: "It's not as if someone has died". To this day, I don't think he realizes how close he came to getting stabbed in the eye with a pipette. Instead of taking bloody revenge, I pointed out that if researchers didn't care so much, he would still be treating head colds with leeches — a less satisfying but more legal response.

Unfortunately, voicing frustrations to colleagues can be just as futile, and prompt the short and not so sweet response: "That's science". To be sure, cultivating a career in the frighteningly competitive world of research leaves no room for hand-holding or mollycoddling, and I truly believe that principal investigators need full-body elephant-hide transplants to achieve the thick skin required for the job. However, we are all human and

everyone needs some sort of coping mechanism. Losing this mechanism is always a disaster.

In truth, there is no magic answer for how to deal with a disappointment rate of 90%. Some people build up walls to protect themselves, but this can result in suppression of all emotion. And let's be honest: given the hours that scientists work and the wages we earn, it is mostly our passion that keeps us chained to the lab bench. Dulling the rare moments of true toe-tingling excitement when things work and we discover something for the first time would be far too big a sacrifice. But others might have quite a different attitude and feel the disappointment so acutely that it destroys their confidence and paralyses them. Channeling your emotion into something manageable is truly important. I suggest that anyone new to research should find healthy ways to deal with their frustrations. Some people read or play a sport; others go out dancing. My coping mechanism is a large glass of red wine and a fantastic group of friends who put up with my rants, then shut me up with a good dose of perspective and insight.

Humans are social animals, and sometimes solitude enhances the feeling of ineptitude and makes dealing with a disappointment even harder. Maybe next time someone wanders past you gazing forlornly at their lab book with that oh-so-familiar look of puzzlement and frustration, a wee pat on the back and a bit of camaraderie might help. Yes, 'that's science'. But perhaps they'll see that success is possible despite repeated failures.

Lydia Soraya Murray is a PhD student in molecular genetics and cell biology at the University of Glasgow, UK.



"We want to train the tool shapers as well as the people more into applying the tools."

Celia van Gelder

And in September, the University of Michigan Medical School in Ann Arbor established a computational-medicine and bioinformatics department to help attract new faculty members and trainees. In June, Emory University School of Medicine in Atlanta, Georgia, launched a biomedical-informatics department with the goal of combining expertise in imaging, computer science and biology to improve patient care. It will recruit four or five researchers over the next few years. "Our particular strength is training computer scientists who want to transition into biomedical informatics, and bringing them together with clinicians to use informatics to treat disease," says department chair Joel Saltz.

Qualified postdocs remain in demand. "It can be very difficult for individual investigators to hire a postdoc in bioinformatics," says Tom Tullius, interim chair of the bioinformatics programme at Boston University in Massachusetts. He attributes the paucity of candidates in part to efforts over the past several years to build large teams at high-powered institutes — such as the Broad Institute in Cambridge, Massachusetts, or the Wellcome Trust Sanger Institute in Cambridge, UK

— leaving smaller labs struggling to find talent. The growth of training programmes could ease this.

Now sequencing centres won't be the sole providers of data, individual researchers, particularly at medical centres, will have ample data to fuel research and training. "We've passed out of the period of genome projects where

there were amazing public data raining down from the heavens; it's now possible to do exciting work without being associated with data-generating centres," says Lee.

Sczyrba, who begins a junior faculty position in metagenomics at the University of Bielefeld Center for Biotechnology in Germany this autumn, says that unpredictability is what makes the discipline so exciting. "We don't know where we will be in ten years because the technologies and ideas are moving so fast," he says. As Cleaver notes: "Perhaps the best career strategy is to stay flexible and curious." ■

Virginia Gewin is a freelance writer in Portland, Oregon.



BEAU LARK/CORBIS

COLUMN

The human touch

A little empathy goes a long way in the competitive confines of a laboratory, argues **Lydia Soraya Murray**.

As almost every scientist knows, a person's first year in research is an emotional minefield. One minute you're flying high. The next, you're banging your head against the wall, resisting the urge to draw in results with a marker pen. Forget the F-word; in science, it's the O-word that generates dread. I sometimes think that 'optimizing' should be spelled 'r.e.p.e.a.t.e.d.f.a.i.l.u.r.e.s'.

After a stimulating yet often soul-destroying start to my PhD, I have decided that coming to terms with the lows is one of the most important things that you can take away from your first year. Never mind the dreaded literature review; this is unquestionably more important.

Because scientists do incredibly specialized and often misunderstood work, it can be hard for people outside our particular fields to empathize with our attachment to our projects. I have a close friend who is a physician. After several weeks of my hard work culminated in what can be described only as 'diddly squat', my friend offered these consoling words: "It's not as if someone has died". To this day, I don't think he realizes how close he came to getting stabbed in the eye with a pipette. Instead of taking bloody revenge, I pointed out that if researchers didn't care so much, he would still be treating head colds with leeches — a less satisfying but more legal response.

Unfortunately, voicing frustrations to colleagues can be just as futile, and prompt the short and not so sweet response: "That's science". To be sure, cultivating a career in the frighteningly competitive world of research leaves no room for hand-holding or mollycoddling, and I truly believe that principal investigators need full-body elephant-hide transplants to achieve the thick skin required for the job. However, we are all human and

everyone needs some sort of coping mechanism. Losing this mechanism is always a disaster.

In truth, there is no magic answer for how to deal with a disappointment rate of 90%. Some people build up walls to protect themselves, but this can result in suppression of all emotion. And let's be honest: given the hours that scientists work and the wages we earn, it is mostly our passion that keeps us chained to the lab bench. Dulling the rare moments of true toe-tingling excitement when things work and we discover something for the first time would be far too big a sacrifice. But others might have quite a different attitude and feel the disappointment so acutely that it destroys their confidence and paralyses them. Channeling your emotion into something manageable is truly important. I suggest that anyone new to research should find healthy ways to deal with their frustrations. Some people read or play a sport; others go out dancing. My coping mechanism is a large glass of red wine and a fantastic group of friends who put up with my rants, then shut me up with a good dose of perspective and insight.

Humans are social animals, and sometimes solitude enhances the feeling of ineptitude and makes dealing with a disappointment even harder. Maybe next time someone wanders past you gazing forlornly at their lab book with that oh-so-familiar look of puzzlement and frustration, a wee pat on the back and a bit of camaraderie might help. Yes, 'that's science'. But perhaps they'll see that success is possible despite repeated failures.

Lydia Soraya Murray is a PhD student in molecular genetics and cell biology at the University of Glasgow, UK.



"We want to train the tool shapers as well as the people more into applying the tools."

Celia van Gelder

HERE BE MONSTERS

Virtually free.

BY STEPHANIE ZVAN

Karee stuck her head into the windowless lab. “Doc?” Doctor Andrews stared at her screen, chewing a twisted strand of hair.

“Uh, Doc?”

“Oh!” Andrews started. “Is it time already?”

“Actually, I’m running late.”

“Right.” Doc Andrews stood up but kept watching the screen.

“Doc...”

“Sorry, sorry.” She turned. “It’s just that I might finally be there. If the lab in Sweden has duplicated my results, we may finally have a cure, even for the worst cases.”

“Really, Doc? That’s wonderful!” Karee’s voice rang in the concrete hallway. “No more muco-whatsis?”

“No more MPS.” Andrews laughed, a sound of pure joy. “No more sick babies. No more stunted bodies and minds. Just healthy children, beautiful and sound.” She licked her lower lip.

Despite her best effort, Karee twitched a little. She scratched her shoulder to cover it. “That’s great, Doc. Great.” It was wonderful news, but... They arrived at the doctor’s room, none too soon. “Here we are. Have you eaten?”

She often forgot.

“While I was waiting,” Andrews grabbed what looked like long underwear off the back of a chair and headed into the bathroom. A few minutes and some running water later, she was changed and back. She sat on her bed. “All set.”

The helmet always looked uncomfortable to Karee, bulky and claustrophobic, and the relish with which Andrews put it on didn’t make Karee any happier. She waited for Andrews to settle into her special pillow. A light on the helmet indicated everything was synching properly. Small power and data cables clipped onto the pyjamas. Gloves attached to the sleeves completed the outfit.

“You good, Doc?”

“Oh, yes.” Andrews gave a little wriggle of anticipation.

Karee swallowed. No getting used to that. “Good night, then.”

With the last of the inmates ‘shelved and synched’, she signed out using her passkey. Her ward was quiet, with the exception of the occasional low moan. Time to leave her charges to the night staff and rejoin society.

There was no good reason to wash her



hands at the end of her shift, but she always did. Her face too. The cool evening breeze found the spots around her ears she hadn’t quite dried. She shivered but felt much lighter than she had inside.

Then she saw the sign through the fence. Poster paper on a broom handle, it said simply: *Here be monsters*. It must be Thursday.

Theo was alone, as he always was these days. The year before, his wife, Hannah, had waded into the lake and swum out farther than she could swim back.

No one at the facility ever spoke to him, but everyone knew his story. Everyone knew about Hannah, and everyone knew about their son.

Claude was born the year the facility opened. Theo and Hannah weren’t protesting then, but plenty of others were. Sure, the paedophiles and compulsive sadists should be locked up, but using VR to give them what they wanted? Victimless as it was, it still felt wrong.

Karee understood. She did. But study after study had showed that no treatment was effective enough in changing inherently antisocial sexual orientations and that the stigma surrounding them only made people more likely to offend. Involuntary commitment after the fact couldn’t help the victims. Voluntary commitment with the

VR as incentive worked

despite its unpopularity.

The video interviews had at least made it acceptable. Doctor

Andrews had been one of the first interviewed, and Karee had transferred from maximum security to guarding voluntary commitment after seeing the look of relief on Andrews’s face. Here Karee could make a difference.

The difference hadn’t come in time for Claude. When the neighbour who raped and murdered him was found to be one of those organizing the local demonstrations, most of the remaining protests stopped. More people volunteered to be locked up. Claude’s death saved uncounted children, but it destroyed his parents. And when their neighbour was murdered in prison two years later, Claude’s family was left without any target for their anger except Karee’s facility.

So, here it was, Thursday again, and there stood Theo with his sign. He hefted it a little higher when Karee was buzzed through the gate, but he didn’t look at her. He never had, not once in the past five years.

Karee sighed, suddenly tired. Doc Andrews and the others were rewarded for making the world a safer place. They were happy. Why wasn’t she? The kids were taken care of, and so were the... well, the monsters who’d agreed not to threaten them. Everybody was taken care of, in fact, except her and...

Karee took a deep breath. “Hey, Theo?”

Startled, Theo looked at her for once. Karee had dreaded seeing anger or hate on his face, but the blankness there disturbed her more. It said he’d forgotten what his protest was supposed to accomplish. He looked terribly old.

Karee nodded up the street. “You’ve got to be just about done here, right? Want to get a cup of coffee?” Coffee seemed like such a simple, uncomplicated good thing, what she wanted more than anything else in the world right now.

The blankness stared back at her. Then Theo opened his mouth. It worked for a bit, no sound coming out, but confusion was an improvement. Karee realized they must be about the same age.

“Come on.” Karee motioned with her hand. “Coffee shop’s just on the corner. I’ll buy.”

Finally, Theo rested his sign against the fence and nodded. His smile was tentative, but it was a smile. As they walked away, Karee felt herself grinning for the first time in years. ■

Stephanie Zvan is a Minneapolis writer fascinated with the strains that science and society place on one another. She blogs on that and more at *Almost Diamonds*.

➔ **NATURE.COM**
Follow Futures on
Facebook at:
[go.nature.com/mtoodm](https://www.facebook.com/mtoodm)

S-nitrosylation of NADPH oxidase regulates cell death in plant immunity

Byung-Wook Yun¹, Angela Feechan^{1†}, Minghui Yin¹, Noor B. B. Saidi^{1,2}, Thierry Le Bihan³, Manda Yu¹, John W. Moore¹, Jeong-Gu Kang¹, Eunjung Kwon¹, Steven H. Spoel¹, Jacqueline A. Pallas^{4†} & Gary J. Loake¹

Changes in redox status are a conspicuous feature of immune responses in a variety of eukaryotes^{1,2}, but the associated signalling mechanisms are not well understood. In plants, attempted microbial infection triggers the rapid synthesis of nitric oxide^{3,4} and a parallel accumulation of reactive oxygen intermediates, the latter generated by NADPH oxidases related to those responsible for the pathogen-activated respiratory burst in phagocytes⁵. Both nitric oxide and reactive oxygen intermediates have been implicated in controlling the hypersensitive response, a programmed execution of plant cells at sites of attempted infection^{3,5,6}. However, the molecular mechanisms that underpin their function and coordinate their synthesis are unknown. Here we show genetic evidence that increases in cysteine thiols modified using nitric oxide, termed S-nitrosothiols, facilitate the hypersensitive response in the absence of the cell death agonist salicylic acid and the synthesis of reactive oxygen intermediates. Surprisingly, when concentrations of S-nitrosothiols were high, nitric oxide function also governed a negative feedback loop limiting the hypersensitive response, mediated by S-nitrosylation of the NADPH oxidase, AtRBOHD, at Cys 890, abolishing its ability to synthesize reactive oxygen intermediates. Accordingly, mutation of Cys 890 compromised S-nitrosothiol-mediated control of AtRBOHD activity, perturbing the magnitude of cell death development. This cysteine is evolutionarily conserved and specifically S-nitrosylated in both

human and fly NADPH oxidase, suggesting that this mechanism may govern immune responses in both plants and animals.

Complex plants do not possess a nitric oxide synthase structurally related to those found in animals; nevertheless, a number of potential sources for pathogen-triggered nitric oxide synthesis have been described, including nitrate reductase and an arginine-dependent nitric-oxide-synthase-like activity⁷. S-nitrosylation, the addition of a nitric oxide moiety to a reactive cysteine thiol to form an S-nitrosothiol⁸ (SNO), is an important route for nitric oxide bioactivity. In *Arabidopsis*, an S-nitrosoglutathione (GSNO) reductase (AtGSNOR1) governs both the concentrations of GSNO and, indirectly, protein SNOs⁸. We determined the temporal profile of SNO concentrations during the development of hypersensitive response in *atgsnor1-3* and *atgsnor1-1* plants, in which AtGSNOR1 activity is absent or increased, respectively⁸. Thus, SNO concentrations were anticipated to be higher in *atgsnor1-3* and lower in *atgsnor1-1* plants, relative to wild type.

SNO concentrations were also determined in the *NO overproducing 1* (*nox1*) mutant⁹. Such plants were challenged with the bacterial pathogen *Pseudomonas syringae* pv. *tomato* (*Pst*) DC3000 expressing either AvrB or AvrRps4 effector proteins, which are recognized by the resistance (*R*) gene products RPM1 and RPS4, respectively, each a prototypic member of a distinct *R* protein subclass^{10,11}. In each case, SNO concentrations increased over time in all the *Arabidopsis* lines tested (Fig. 1a, b), relative to *Pst* DC3000 controls (Supplementary Fig. 1). However,

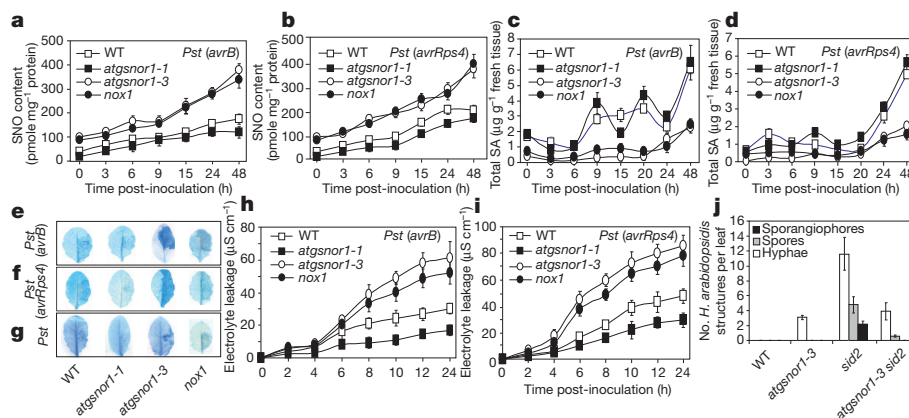


Figure 1 | SNOs positively regulate cell death by hypersensitive response. **a**, Profile of SNO accumulation following challenge with *Pst* DC3000 (*avrB*). **b**, SNO accumulation following attempted *Pst* DC3000 (*avrRps4*) infection. **c**, Total salicylic acid (SA) accumulation in response to attempted *Pst* DC3000 (*avrB*) colonization. **d**, Accumulation of salicylic acid following *Pst* DC3000 (*avrRps4*) challenge. **e**, Cell death development in the given *Arabidopsis* genotypes, triggered by 5×10^6 c.f.u. ml⁻¹ *Pst* DC3000 (*avrB*) at 15 h.p.i. and scored by trypan blue staining. **f**, Magnitude of cell death development in the stated plant lines following challenge with 5×10^6 c.f.u. ml⁻¹ *Pst* DC3000 (*avrRps4*) at 15 h.p.i.,

determined by trypan blue staining. **g**, Cell death development in the given *Arabidopsis* genotypes, triggered by 5×10^6 c.f.u. ml⁻¹ *Pst* DC3000 at 15 h.p.i. and scored by trypan blue staining. **h**, Extent of cell death development established by electrolyte leakage in the given *Arabidopsis* genotypes following challenge with *Pst* DC3000 (*avrB*) at 1×10^8 c.f.u. ml⁻¹. **i**, Quantification of cell death by electrolyte leakage in the stated plant lines in response to *Pst* DC3000 (*avrRps4*). **j**, Growth of *H. arabidopsidis* Emwa1 at 10 d post-inoculation in the given *Arabidopsis* genotypes. Data points represent mean \pm s.e. ($n = 3$). Unless stated otherwise, avirulent strains of *Pst* DC3000 were infiltrated at 1×10^6 c.f.u. ml⁻¹.

¹Institute of Molecular Plant Sciences, School of Biological Sciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3JH, UK. ²University of Putra, Department of Cell and Molecular Biology, Serdang 43400, Malaysia. ³University of Edinburgh, Centre for Systems Biology, Edinburgh EH9 3JR, UK. ⁴Trait Research, Syngenta, Jealott's Hill, Bracknell, Berkshire RG42 6EY, UK. [†]Present addresses: CSIRO Plant Industry, PO Box 350, Glen Osmond, South Australia 5064, Australia (A.F.); Bloomsbury Centre for Bioinformatics, University College London, Gower Street, London WC1E 6BT, UK (J.A.P.).

concentrations of these molecules were higher in *atgsnor1-3* and *nox1* plants than in wild type. Similar results were obtained when we used reporters of NO accumulation¹² (Supplementary Fig. 2a–d).

We also determined the profile of salicylic acid accumulation (salicylic acid is a cell death agonist¹³) in these lines. Total salicylic acid accumulation was diminished in *atgsnor1-3* and *nox1* plants relative to wild type (Fig. 1c, d and Supplementary Fig. 3a), as was free salicylic acid and salicylic acid β -glucoside (Supplementary Fig. 3b–e). Together, these results suggest that *atgsnor1-3* and *nox1* plants accrue markedly more SNOs over time during the development of hypersensitive response, and that the *atgsnor1-1* line accumulates significantly fewer. Further, salicylic acid concentrations are diminished in *atgsnor1-3* and *nox1* plants.

Next we assessed the development of hypersensitive response in these plants. Challenge with *Pst* DC3000 expressing either *avrB* or *avrRps4* revealed that this defence response was delayed in *atgsnor1-1* plants relative to wild type. In contrast, the development of hypersensitive response in *atgsnor1-3* and *nox1* plants was accelerated (Supplementary Fig. 4a, b). To determine the extent of cell death by hypersensitive response (CDHR), a smaller inoculum of *Pst* DC3000 strains was used and the resulting leaves were stained with trypan blue, which marks dead or dying plant cells⁵. Relative to wild type, *atgsnor1-3* and *nox1* plants showed a prominent increase in cell death, but this response was markedly reduced in the *atgsnor1-1* line (Fig. 1e–g). We corroborated these findings by quantifying cell-death-induced electrolyte leakage. Again, cell death was significantly greater in *atgsnor1-3* and *nox1* plants than in wild type, and there was a decrease in the *atgsnor1-1* line (Fig. 1h, i and Supplementary Fig. 5). To confirm and extend these findings, we studied the effect of high SNO concentrations mediated by *atgsnor1-3* on the hypersensitive response in the absence of SALICYLIC ACID INDUCTION DEFICIENT 2 (*SID2*) function, which is required for pathogen-triggered salicylic acid synthesis¹⁴. There was no significant difference in the extent of cell death development in the *atgsnor1-3 sid2* double mutant relative to the *atgsnor1-3* line (Supplementary Fig. 6). Collectively, these findings imply that in *atgsnor1-3* and *nox1* plants, the development of CDHR has accelerated kinetics and increased magnitude. However, in *atgsnor1-1* plants, cell death development is reduced. Thus, despite diminished salicylic acid concentrations, a greater concentration of SNO positively regulates the development of the hypersensitive response mediated by at least two distinct *R* gene subclasses.

CDHR does not seem to be required for limiting bacterial infection^{5,15}. Therefore, to identify a potential role for SNO-driven cell death in disease resistance, we challenged *atgsnor1-3* plants with the avirulent oomycete *Hyaloperonospora arabidopsidis* isolate Emwa1, which is recognized by *RPP4* (ref. 16). The death of challenged host cells has been proposed as a key resistance mechanism against oomycetes^{5,17}. As expected, cell death was more pronounced in *atgsnor1-3* plants, relative to wild type, in response to Emwa1 (Supplementary Fig. 7). In addition to its role in cell death, salicylic acid is also a key immune activator, and plants defective in its accumulation routinely show diminished defence responses¹⁴. Indeed, the *Arabidopsis sid2* mutant, in which pathogen-induced salicylic acid accumulation is reduced^{14,18}, was compromised in *RPP4*-mediated resistance against Emwa1 (Fig. 1j). By contrast, even though salicylic acid concentrations were equally reduced, relative to wild type, in *atgsnor1-3* plants (Supplementary Fig. 8a–c), Emwa1 failed to complete its life cycle in these plants in the same way it did in the resistant wild-type line (Fig. 1j). An *atgsnor1-3 sid2* double mutant also had increased resistance against Emwa1 relative to *sid2* plants (Fig. 1j). Moreover, although SNO and nitrite concentrations are higher in *atgsnor1-3* plants, in *sid* mutants they are comparable to wild type (Supplementary Fig. 9a, b). Therefore, cell death development mediated by increased SNO is sufficient to convey resistance against Emwa1 in the absence of salicylic acid accumulation and associated defence responses.

NO function is thought to be closely interconnected with that of reactive oxygen intermediates (ROIs) in cell death development^{3,6}. We therefore monitored ROI accumulation activated by distinct *R* proteins in *atgsnor1* and *nox1* mutants, as determined by 3',3'-diaminobenzidine (DAB) staining⁵. Relative to wild type, *atgsnor1-3* and *nox1* plants showed decreased pathogen-induced ROI accumulation, whereas mutant *atgsnor1-1* plants accumulated more ROIs even in the absence of pathogen challenge (Fig. 2a, b and Supplementary Fig. 10a, b). Hence, in addition to being autonomous of salicylic acid, increased SNO concentrations may also facilitate CDHR independently of DAB-detectable ROI accumulation. To explore this possibility further, we studied the effect of high SNO concentrations on the hypersensitive response in the absence of the NADPH-dependent oxidases AtRBOHD and AtRBOHF, which drive pathogen-induced ROI synthesis⁵. As expected, *atrbohD* and *atrbohF* single and double mutants showed decreased pathogen-induced CDHR than did wild-type plants, indicating that ROI synthesis is required for full development of the hypersensitive response (Fig. 2c and Supplementary Fig. 11). However, CDHR was not significantly different in *atgsnor1-3 atrbohD*, *atgsnor1-3 atrbohF* and *atgsnor1-3 atrbohD atrbohF* mutants than in *atgsnor1-3* plants (Fig. 2c and Supplementary Figs 11 and 12a), despite the reduced ROI accumulation in these lines (Supplementary Fig. 12b). These results suggest that high SNO concentrations can facilitate CDHR independently of ROI synthesis mediated by AtRBOHD or AtRBOHF.

Next we asked how SNO concentrations governed by AtGSNOR1 are able to manipulate ROI concentrations. AtRBOHD is responsible for virtually all the DAB-detectable ROIs induced by avirulent strains of *Pst* DC3000⁵. Thus, it is plausible that SNO concentrations regulate AtRBOHD function. However, basal or pathogen-triggered changes in SNO concentration did not influence AtRBOHD protein concentrations, as determined by monitoring a Myc–AtRBOHD fusion protein detected by a Myc-specific antibody in *atrbohD atgsnor1* double

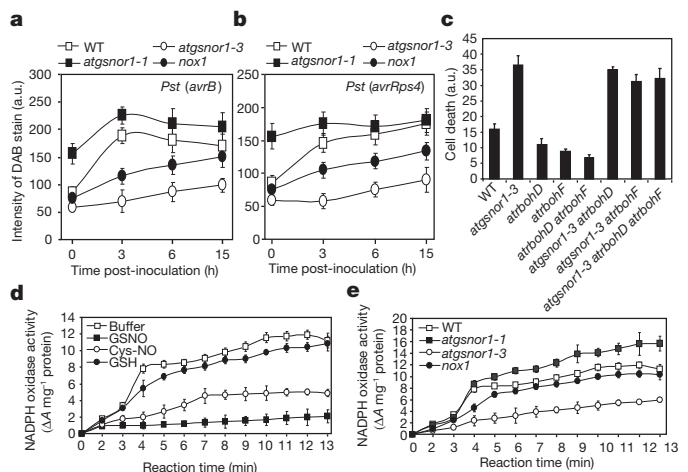


Figure 2 | Increased SNO concentrations blunt NADPH oxidase activity and reduce ROI accumulation. **a**, ROI accumulation determined by DAB staining in the given *Arabidopsis* genotypes following challenge with *Pst* DC3000 (*avrB*). a.u., arbitrary units. **b**, Accumulation of ROIs in the stated plant lines in response to attempted *Pst* DC3000 (*avrRps4*) infection. **c**, Cell death development in *atgsnor1* double and triple mutants in response to attempted *Pst* DC3000 (*avrB*) infection at 1×10^6 c.f.u. ml⁻¹ at 48 h.p.i. A Student's *t*-test comparing cell death in *atgsnor1-3* with that in *atgsnor1-3 atrbohD* plants, in a similar fashion to the other double and triple mutants, established that there was no statistically significant difference ($P = 0.6466$). **d**, NADPH oxidase activity in *Arabidopsis* following exposure to the given natural nitric oxide donors or related control treatments. ΔA at 480 nm, change in absorbance. **e**, NADPH oxidase activity in given *Arabidopsis* lines at 24 h.p.i. following challenge with *Pst* DC3000 (*avrB*). Data points represent mean \pm s.e. ($n = 3$). Avirulent strains of *Pst* DC3000 were infiltrated at 1×10^6 c.f.u. ml⁻¹ unless stated otherwise.

mutants (Supplementary Fig. 13). To explore whether SNOs could directly regulate NADPH oxidase activity, microsomal preparations from pathogen-challenged wild-type leaves were treated with the natural nitric oxide donor GSNO or *S*-nitroso-L-cysteine (Cys-NO), and NADPH oxidase activity was determined. Exposure to either GSNO or, to a lesser extent, Cys-NO significantly reduced the activity of this enzyme relative to the buffer control treatment (Fig. 2d). Furthermore, the absence of an effect following exposure to reduced glutathione (GSH) confirmed the specificity of this response. To determine the possible biological consequences of these findings, we measured the activity of this protein in *atgsnor1-3*, *atgsnor1-1*, *nox1* and wild-type leaves challenged with *Pst* DC3000 (*avrB*). NADPH oxidase activity was significantly reduced in *atgsnor1-3* and *nox1* plants that have high SNO concentrations (Fig. 2e). Collectively, these findings suggest that changes in SNO concentrations can modulate NADPH oxidase activity, implying that this protein might be regulated by *S*-nitrosylation.

To determine whether NADPH oxidase might be *S*-nitrosylated, recombinant protein was exposed to a range of either GSNO or Cys-NO concentrations typically used to score for *S*-nitrosylation *in vitro*^{19–21}, and was monitored for the formation of SNO–AtRBOHD by the biotin-switch method²². AtRBOHD was *S*-nitrosylated in a concentration-dependent fashion by either GSNO or Cys-NO (Fig. 3a, b). Furthermore, the addition of dithiothreitol strikingly reduced the concentration of SNO–AtRBOHD formation (Fig. 3b), consistent with the presence of a reversible thiol modification. AtRBOHD contains a number of cysteines that might serve as sites for this redox-based modification. The carboxy- and amino-terminal portions of this protein were therefore expressed, exposed to GSNO and subjected to biotin-switch analysis. Only the C-terminal portion of AtRBOHD was *S*-nitrosylated (Fig. 3c). The C-terminal portion of AtRBOHD has two cysteines: Cys 825 and Cys 890. These residues were therefore mutated either individually or in combination, and the resulting proteins were expressed, treated with GSNO and analysed by the biotin-switch method. The Cys890Ala mutation but not the Cys825Ala mutation abolished *S*-nitrosylation of AtRBOHD (Fig. 3d). Mass spectrometry analysis was also consistent with *S*-nitrosothiol formation at Cys 890 (Supplementary Fig. 14). Collectively, these findings imply that AtRBOHD is specifically *S*-nitrosylated *in vitro* on Cys 890.

Cys 890 is evolutionarily conserved, suggesting that NADPH oxidases from other eukaryotes might also be *S*-nitrosylated (Supplementary Fig. 15). We therefore exposed recombinant human and *Drosophila* RBOH proteins to either GSNO or Cys-NO. Both of these proteins were specifically *S*-nitrosylated (Supplementary Fig. 16a, b). The site of SNO formation was Cys 537 for human NOX2 (also known as CYBB) and Cys 1315 for *Drosophila* Nox (Supplementary Fig. 16c,

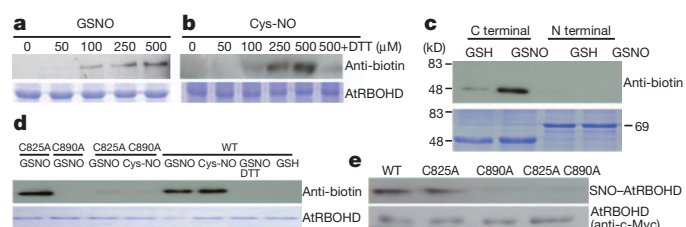


Figure 3 | *S*-nitrosylation of AtRBOHD. **a**, *S*-nitrosylation of recombinant AtRBOHD *in vitro* by the nitric oxide donor GSNO. **b**, Cys-NO *S*-nitrosylates recombinant AtRBOHD and this is reversible by treatment with dithiothreitol (DTT). **c**, The C terminus of AtRBOHD is *S*-nitrosylated by GSNO. **d**, *S*-nitrosylation analysis of wild-type and mutant derivatives of AtRBOHD. **e**, *In vivo* *S*-nitrosylation of wild-type and mutant derivatives of AtRBOHD in the given plant genotypes 24 h after 5×10^6 c.f.u. ml⁻¹ *Pst* DC3000 (*avrB*) infiltration. AtRBOHD was detected by virtue of its c-Myc tag using antibodies raised against this tag. AtRBOH proteins in **a–d** were examined 20 min post-exposure to nitric oxide donors. All experiments were repeated at least twice, with similar results.

d), both of which corresponded to Cys 890 of AtRBOHD (Supplementary Fig. 15). Together, these data suggest that NADPH oxidases from at least two animals are specifically *S*-nitrosylated at this conserved cysteine, raising the possibility that this redox modification might regulate the activity of these enzymes in many other eukaryotes. Interestingly, an organizer protein, absent from plants, that interacts with animal NADPH oxidases may also be subject to *S*-nitrosylation in endothelial cells²³. Therefore, in mammals NADPH oxidase function might also be regulated indirectly by another reactive cysteine residue.

To determine whether AtRBOHD is *S*-nitrosylated *in vivo* during the hypersensitive response, transgenic *atrbohD* lines expressing either Myc-tagged wild-type AtRBOHD or mutant derivatives were challenged with *Pst* DC3000 (*avrB*). Subsequently, endogenous proteins were subjected to biotin-switch analysis and biotinylated proteins purified with streptavidin beads. These proteins were then immunoblotted with an anti-Myc antibody. Both wild-type AtRBOHD and the Cys825Ala mutant were *S*-nitrosylated, but the Cys890Ala mutant and the Cys825Ala Cys890Ala double mutant were not (Fig. 3e). Thus, AtRBOHD is specifically *S*-nitrosylated *in vivo* at Cys 890 during the plant defence response.

To understand whether *S*-nitrosylation of Cys 890 can modulate the activity of AtRBOHD, we first computationally modelled the structure of this protein. This indicated that Cys 890 is positioned closely behind the conserved Phe 921 and Phe 570 residues in AtRBOHD and NOX2, respectively. Similar to the homologous Tyr 247 in ferredoxin reductase²⁴, these residues are expected to have a significant role in binding flavin adenine dinucleotide (FAD). Accordingly, mutation of Phe 570 was reported to impair NOX2 function²⁵. The model further predicts that *S*-nitrosylation of AtRBOHD at Cys 890 may disrupt the side-chain position of Phe 921, impeding FAD binding (Supplementary Fig. 17a, b). To extend these findings, we determined the consequences of introducing an *S*-nitrosylated Cys 890 into our model. This disrupted the coplanar localization of Phe 921, thereby destabilizing or sterically ejecting FAD (Supplementary Fig. 18a, b). Consistent with these predictions, we found that prior GSNO exposure markedly reduced the binding of this cofactor to AtRBOHD. Conversely, GSNO did not diminish FAD binding in the Cys 890 AtRBOHD mutant (Fig. 4a).

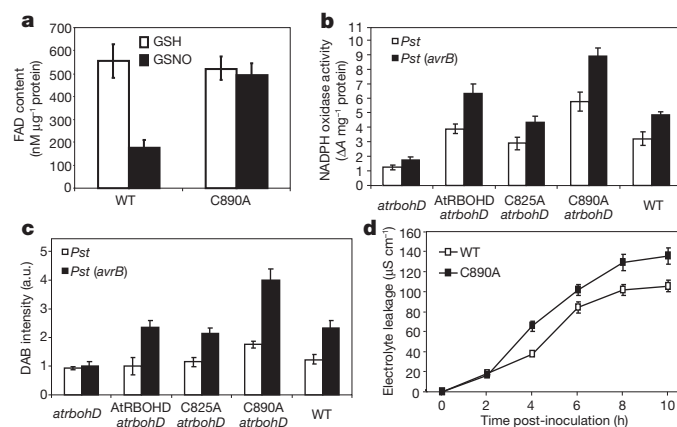


Figure 4 | The AtRBOHD Cys890Ala mutant shows increased activity during the defence response, amplifying ROI accumulation and cell death development. **a**, Quantification of FAD binding following prior GSNO or GSH treatment of wild-type AtRBOHD or the mutant Cys890Ala derivative. **b**, NADPH oxidase activity in plant extracts from the given *Arabidopsis* genotypes infiltrated with *Pst* DC3000 (white) or *Pst* DC3000 (*avrB*) (black) at 24 h.p.i. **c**, ROI accumulation determined by DAB staining in the leaves of given plant lines infiltrated with *Pst* DC3000 (white) and *Pst* DC3000 (*avrB*) (black) at 24 h.p.i. **d**, Extent of cell death development determined by electrolyte leakage in the given wild-type and mutant *Arabidopsis* lines following challenge with *Pst* DC3000 (*avrB*) infiltrated at 1×10^8 c.f.u. ml⁻¹. Data points represent mean \pm s.e. ($n = 3$). Unless stated otherwise, all strains of *Pst* DC3000 were infiltrated at 1×10^6 c.f.u. ml⁻¹.

Thus, S-nitrosylation of AtRBOHD may preclude FAD binding inhibiting the activity of this enzyme.

To explore the possible biological consequence of AtRBOHD SNO formation at Cys 890 and the resulting loss of FAD binding, we monitored NADPH oxidase activity in mutant *atrbohD* plants expressing either a wild-type AtRBOHD transgene or the Cys890Ala derivative. Pathogen-induced levels of leaf NADPH oxidase activity were significantly increased in the Cys890Ala mutant, relative to wild type (Fig. 4b). Furthermore, pathogen-induced ROI accumulation was greater in the Cys890Ala mutant line than in wild type (Fig. 4c). These data imply that S-nitrosylation of AtRBOHD at Cys 890 during the development of hypersensitive response blunts NADPH oxidase activity, resulting in decreased ROI generation. To determine the possible impact of the AtRBOHD Cys890Ala mutation on cell death, we challenged plants possessing this mutation with *Pst* DC3000 (*avrB*) and scored the extent of their response. There was a significant increase in cell death in the AtRBOHD Cys890Ala line relative to wild type (Fig. 4d). Thus, enhanced ROI production in the Cys890Ala mutant, which is not a target for S-nitrosylation, facilitates CDHR. This work was confirmed and extended by determining in parallel SNO accumulation, NADPH oxidase activity and cell death development in the AtRBOHD Cys890Ala mutant line, relative to wild type, following attempted pathogen ingress (Supplementary Fig. 19a–c). Together, this information implies that S-nitrosylation of AtRBOHD at Cys 890 promoted by increasing SNO concentrations serves to curb excessive cell death by blunting AtRBOHD-dependent ROI synthesis.

These data establish a molecular framework for SNO function during the hypersensitive response. We speculate that on attempted infection, total cellular SNOs governed by AtGSNOR1 and ROIs synthesized by AtRBOHD⁵, in combination with salicylic acid accumulation¹³, positively regulate the development of cell death. However, as SNO concentrations increase during the pathogen-triggered nitrosative burst, salicylic acid accumulation is reduced and further S-nitrosylation of AtRBOHD at Cys 890 decreases ROI synthesis. Collectively, this molecular dialogue may serve to limit cell death development during the hypersensitive response (Supplementary Fig. 20).

Our findings are reminiscent of those in animals, where increasing concentrations of total cellular SNOs drive apoptosis in a variety of tissues¹². Conversely, S-nitrosylation of the pro-apoptotic regulators nuclear factor κ B²⁶ and caspase cysteine protease²⁷ function to constrain cell death development. Furthermore, SNO formation at a conserved cysteine in NADPH oxidases from plants, humans and flies implies that this novel pro-survival mechanism might operate during immune function throughout such complex eukaryotes.

METHODS SUMMARY

Pathogen inoculations and small-molecule measurements. A variety of *Pst* DC3000 strains and *H. arabidopsidis* isolates were used for pathogen inoculations, as indicated. Trypan blue and DAB staining were conducted for detection of cell death and ROI accumulation, respectively²⁸. Salicylic acid was determined using a mini-scale procedure²⁹ based on high-pressure liquid chromatography. SNO contents were analysed using a chemiluminescence-based procedure³⁰. Diaminofluorescein diacetate imaging was applied for the detection of nitric oxide. Electrolyte leakage was measured by using a DiST WP conductivity meter to quantify cell death.

Gene constructs and transgenic plants. For cloning of NADPH oxidase genes and site-directed mutagenesis, the pGEX4T-1 vector and the QuickChange II Site-Directed Mutagenesis Kit (Stratagene) were used, respectively. Transgenes were generated by using the pUNI51 cloning vector and the pKYLX-myc9-loxp binary vector, and subsequently were introduced into *Arabidopsis* plants. See Methods for full details.

Biochemistry and computational modelling. NADPH oxidase activity was determined by using epinephrine and NADPH as substrates. *In vitro* and *vivo* S-nitrosylation assays used either an anti-biotin or an anti-Myc9 antibody, respectively. Mass spectrometry was done using a capillary gas high-pressure liquid chromatography tandem mass spectrometry analysis system. For computational

modelling of NOX2 and AtRBOHD, we used Phyre, DeepView, SwissModel and Jpred as described in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 May 2010; accepted 4 August 2011.

Published online 2 October 2011.

- MacMicking, J. D. *et al.* Identification of nitric oxide synthase as a protective locus against tuberculosis. *Proc. Natl Acad. Sci. USA* **94**, 5243–5248 (1997).
- Tada, Y. *et al.* Plant immunity requires conformational changes of NPR1 via S-nitrosylation and thioredoxins. *Science* **321**, 952–956 (2008).
- Delledonne, M., Xia, Y., Dixon, R. A. & Lamb, C. J. Nitric oxide functions as signal in plant disease resistance. *Nature* **394**, 585–588 (1998).
- Durner, J., Wendehenne, D. & Klessig, D. F. Defense gene induction in tobacco by nitric oxide, cyclic GMP and cyclic ADP ribose. *Proc. Natl Acad. Sci. USA* **95**, 10328–10333 (1998).
- Torres, M. A., Dangel, J. L. & Jones, J. D. *Arabidopsis* gp91^{phox} homologues *AtrbohD* and *AtrbohF* are required for accumulation of reactive oxygen intermediates in the plant defense response. *Proc. Natl Acad. Sci. USA* **99**, 517–522 (2002).
- Delledonne, M., Zeier, J., Marocco, A. & Lamb, C. Signal interactions between nitric oxide and reactive oxygen intermediates in the plant hypersensitive disease resistance response. *Proc. Natl Acad. Sci. USA* **98**, 13454–13459 (2001).
- Leitner, M., Vandelle, E., Gaupels, F., Bellin, D. & Delledonne, M. NO signals in the haze: nitric oxide signalling in plant defence. *Curr. Opin. Plant Biol.* **12**, 451–458 (2009).
- Feechan, A. *et al.* A central role for S-nitrosothiols in plant disease resistance. *Proc. Natl Acad. Sci. USA* **102**, 8054–8059 (2005).
- He, Y. *et al.* Nitric oxide represses the *Arabidopsis* floral transition. *Science* **305**, 1968–1971 (2004).
- Grant, M. R. *et al.* Structure of the *Arabidopsis* RPM1 gene enabling dual specificity disease resistance. *Science* **269**, 843–846 (1995).
- Gassmann, W., Hinsch, M. E. & Staskawicz, B. J. The *Arabidopsis* RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease resistance genes. *Plant J.* **20**, 265–277 (1999).
- Liu, L. *et al.* Essential roles of S-nitrosothiols in vascular homeostasis and endotoxic shock. *Cell* **116**, 617–628 (2004).
- Shirasu, K., Nakajima, H., Rajasekhar, V. K., Dixon, R. A. & Lamb, C. J. Salicylic acid potentiates an agonist-dependent gain control that amplifies pathogen signals in the activation of defense mechanisms. *Plant Cell* **9**, 261–270 (1997).
- Wildermuth, M. C., Dewdney, J., Wu, G. & Ausubel, F. M. Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* **414**, 562–565 (2001).
- Yu, I. C., Parker, J. & Bent, A. F. Gene-for-gene disease resistance without the hypersensitive response in *Arabidopsis* *dnd1* mutant. *Proc. Natl Acad. Sci. USA* **95**, 7819–7824 (1998).
- Holub, E. B., Beynon, J. L. & Crute, I. R. Phenotypic and genotypic characterization of interactions between isolates of *Peronospora parasitica* and accessions of *Arabidopsis thaliana*. *Mol. Plant Microbe Interact.* **7**, 223–239 (1994).
- Keller, H. *et al.* Pathogen-induced elicitor production in transgenic tobacco generates a hypersensitive response and nonspecific disease resistance. *Plant Cell* **11**, 223–236 (1999).
- Nawrath, C. & Metraux, J. P. Salicylic acid induction-deficient mutants of *Arabidopsis* express *PR-2* and *PR-5* and accumulate high levels of camalexin after pathogen inoculation. *Plant Cell* **11**, 1393–1404 (1999).
- Wang, Y.-Q. *et al.* S-nitrosylation of ATSBP3 antagonises the expression of plant immunity. *J. Biol. Chem.* **284**, 2131–2137 (2009).
- Romero-Puertas, M. C. *et al.* S-nitrosylation of peroxiredoxin II E promotes peroxynitrite-mediated tyrosine nitration. *Plant Cell* **19**, 4120–4130 (2007).
- Lindermayr, C., Sell, S., Müller, B., Leister, D. & Durner, J. Redox regulation of the NPR1-TGA1 system of *Arabidopsis thaliana* by nitric oxide. *Plant Cell* **22**, 2894–2907 (2010).
- Jaffrey, S. R., Erdjument-Bromage, H., Ferris, C. D., Tempst, P. & Snyder, S. H. Protein S-nitrosylation: a physiological signal for neuronal nitric oxide. *Nature Cell Biol.* **3**, 193–197 (2001).
- Selemidis, S., Disting, G. J., Peshavariya, H., Kemp-Harper, B. K. & Drummond, G. R. Nitric oxide suppresses NADPH oxidase-dependent superoxide production by S-nitrosylation in human endothelial cells. *Cardiovasc. Res.* **75**, 349–358 (2007).
- Ingelman, M., Bianchi, V. & Eklund, H. The three-dimensional structure of flavodoxin reductase from *Escherichia coli* at 1.7 Å resolution. *J. Mol. Biol.* **268**, 147–157 (1997).
- Zhen, L., Yu, L. & Dinanier, M. C. Probing the role of the carboxyl terminus of the gp91^{phox} subunit of neutrophil flavocytochrome b₅₅₈ using site-directed mutagenesis. *J. Biol. Chem.* **273**, 6575–6581 (1998).
- Matthews, J. R. *et al.* Inhibition of NF- κ B DNA binding by nitric oxide. *Nucleic Acids Res.* **24**, 2236–2242 (1996).
- Mannick, J. B. *et al.* Fas-induced caspase denitrosylation. *Science* **284**, 651–654 (1999).
- Yun, B.-W. *et al.* Loss of actin cytoskeletal function and EDS1 activity, in combination, severely compromises non-host resistance in *Arabidopsis* against wheat powdery mildew. *Plant J.* **34**, 768–777 (2003).
- Aboul-Soud, M. A. M., Cook, K. & Loake, G. J. Measurement of salicylic acid by a high-performance liquid chromatography procedure based on ion-exchange. *Chromatographia* **59**, 129–133 (2004).

30. Liu, L. *et al.* Essential roles of S-nitrosothiols in vascular homeostasis and endotoxic shock. *Cell* **116**, 617–628 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to acknowledge R. Innes and W. Gassmann for *Pst* DC3000 strains expressing either *avrB* or *avrRps4*, respectively. *Arabidopsis* transfer DNA insertion mutants were obtained from SAIL (Syngeneta) populations. We thank M. Tör for the *H. arabidopsidis* isolate Emwa1, and K. Kanchanawatee for the *Drosophila* cDNA clone and associated mutant. A.F. was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) CASE studentship. B.-W.Y. and E.K. were funded by BBSRC grant BB/D011809/1 awarded to G.J.L. J.W.M. received a grant from the Physical Sciences Research Council (EPSRC). M. Yu was the recipient of a Darwin Trust Scholarship. T.L.B. was supported by BBSRC and EPSRC grant BB/D019621/1.

N.B.B.S. and M. Yin were funded by a Ministry of Education Malaysia scholarship and a Torrance Scholarship, respectively.

Author Contributions G.J.L. designed the research and wrote the paper, and with B.-W.Y. planned experiments and analyses. B.-W.Y. conducted the majority of experiments. A.F., M. Yin, N.B.B.S., J.-G.K., J.W.M., M. Yu, E.K. and T.L.B. conducted experiments. S.H.S. generated and interrogated three-dimensional models. J.A.P. was the industrial supervisor of A.F. All authors, especially B.-W.Y. and S.S., commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.J.L. (gloake@ed.ac.uk).

METHODS

Histochemical analysis and small-molecule measurements. Cells committed to die were visualized with lactophenol-trypan blue as described previously²⁸. Peroxides were stained with DAB²⁸. Catalase effectively eliminated DAB staining. H₂O₂-dependent DAB staining was observed in all the *Arabidopsis* lines described in this work. For quantification of either cell death or H₂O₂ accumulation, the extent of either trypan blue or DAB staining, respectively, was determined in six leaves per line in arbitrary units, for each time point and from three independent experiments, using the 'saturation' function in PaintShop Pro 8 (Corel).

Salicylic acid and salicylic acid β -glucoside concentrations were determined using a mini-scale procedure²⁹ based on high-pressure liquid chromatography. For the identification of SNOs, tissue extracts were generated at given times after pathogen inoculation and these samples were analysed by a chemiluminescence-based procedure (Sievers nitric oxide analyser) as described previously³⁰.

For diaminofluorescein diacetate imaging, rosette leaves were incubated for 15 min in a solution of 15 μ M diaminofluorescein diacetate (Alexis) containing 5 mM MES KOH, pH 5.7, 0.25 mM KCl and 1 mM CaCl₂, and were then washed for 5 min (ref. 31). Fluorescent signals were detected using a Bio-Rad Radiance 2100 confocal microscope (Nikon Eclipse, TE2000-U). The dye was excited at 488 nm, and images were collected at emission wavelengths of 500–530 nm. Green fluorescence-specific intensities were quantified using IMAGEJ (version 1.45h).

Plant material and pathogen inoculations. *Arabidopsis* accession Col-0 and cognate mutants were grown under 16 h of light at 22 °C and 8 h of darkness at 18 °C. *Pst* DC3000 was grown, maintained and inoculated as described previously³². *Pst* DC3000 strains were inoculated at the concentrations stated. *H. arabidopsidis* infections were as described previously¹⁶.

Electrolyte leakage. The protocol for electrolyte leakage was adapted from a previously described method³³. Four-week-old plants were injected with bacteria in 10 mM MgCl₂. Ten minutes after injection, 5.0-mm-diameter leaf disks were collected from the injected area and washed extensively with water for 10 min, and then ten discs were placed in a Petri dish with 6 ml of water. Conductivity measurements (three replicates for each treatment) were taken over time by using a DiST WP conductivity meter (HANNA Instruments). The units of this measurement are microsiemens per centimetre, where the distance refers to that between the electrodes.

Cloning of NADPH oxidase genes and site-directed mutagenesis. The primers used to clone the N terminus (from K2 to N357; 356 amino acids) and the C terminus (from K756 to P921; 166 amino acids) of AtRBOHD, the C terminus (from N430 to E540; 111 amino acids) of *Homo sapiens* NOX2 (gi_163854302) and the C terminus (from I1198 to E1338; 141 amino acids) of *Drosophila melanogaster* Nox (gi_161077139) are as follows. AtRBOHD N terminus: 5'-AAGGATCCAAATGAGACGAGGCAATTCA-3' (forward); 5'-TTCTAGTTGCTCTTTTGGCGGTCT-3' (reverse). AtRBOHD C terminus: 5'-AAGGATCCAAAGGACATCATCAACAACATG-3' (forward); 5'-TTCTAGAAGTTCTCTTGTGGAA-3' (reverse). NOX2 C terminus: 5'-GTAATGGATCCAAACGCCCAATCT-3' (forward); 5'-TATGCTCTCGAGTCATTCAGGTCCACAGA-3' (reverse). Nox C terminus: 5'-GAAGAGCAAAAAGCGGAGTC-3' (forward); 5'-GGATTGCGTTTCGTGAAGA-3' (reverse). The amplified PCR products were cloned into pGEX4T-1 vector at the sites of BamHI/EcoRI, BamHI/XhoI and EcoRI/XhoI, respectively.

Site-directed mutagenesis was carried out with QuickChange II Site-Directed Mutagenesis Kit (Stratagene). All procedures followed the manufacturer's manual and specific primers used for AtRBOHD, NOX2 and Nox mutations are as follows. AtRBOHD C825A: 5'-GAGCTTCACAATTATGCCACGAGTGTGTACGA-3' (forward); 5'-TCGTACACACTCGTGGCATAATTGTGAAGCTC-3' (reverse). AtRBOHD C890A: 5'-ATAGGAGTCTTCTACGCTGGAATGCCAGGAAT-3' (forward); 5'-ATTCCTGGCATTCAGCGTAGAAGACTCCTAT-3' (reverse). NOX2 C537A: 5'-ATAGGAGTTTCTCTGCGGACCTGAATGACTC-3' (forward); 5'-GAGTCATTCAGGTCCGCGGAGGAAAACCTCTAT-3' (reverse). Nox C1315A: 5'-GTCACCGTCTTCTACGCCGCCGCCACACAGTTG-3' (forward); 5'-CAACTGTGGTGGGCCGCGGTAGAAGACGGTGAC-3' (reverse). Underlining denotes the codon of alanine.

Transgenic plant materials. The *atrbohD* line was complemented with a wild-type copy of AtRBOHD and also transformed with C825A, C890A or the double mutant. Briefly, a wild-type copy and a mutant copy of AtRBOHD were amplified by using forward primer 5'-CGGAATTCGGATGAAAATGAGACGAGGCAATTCA-3' and reverse primer 5'-CGGGATCCTAGAAGTTCTCTTTGTGGAAGTC-3', and cloned into pUNI51 vector (EcoRI/BamHI). Subsequently, the pUNI51 constructs harbouring AtRBOHD were recombined with pKYLX-myc9-loxP binary vector by a Cre recombinase³⁴. The resulting constructs were introduced into *Agrobacterium* strain GV3101 and subsequently transformed into *Arabidopsis* plants. Transgenic T1 plants were identified by kanamycin selection.

NADPH oxidase biochemistry. NADPH oxidase activity was measured as described previously³⁵. Briefly, 1 g of leaf tissue was ground in liquid nitrogen and dissolved in extraction buffer (0.25 M sucrose, 50 mM HEPES, pH 7.2, 3 mM EDTA, 1 mM dithiothreitol, 0.6% PVP, 3.6 mM L-cysteine, 0.1 mM MgCl₂ and protease inhibitor tablet (Roche)). The crude extract was centrifuged at 10,000g for 30 min and the resulting supernatant was ultracentrifuged at 203,000g for 1 h. The resulting pellet was resuspended in extraction buffer and used as the membrane fraction to measure NADPH oxidase activity spectrophotometrically at 480 nm using epinephrine and NADPH as substrates.

Expression and purification of recombinant proteins. Recombinant proteins were expressed in *Escherichia coli* strain BL21 (DE3) by adding 0.3 mM IPTG with a 6-h incubation. The GST-tagged proteins were purified using a MagneGST Protein Purification System (Promega). All procedures followed the manufacturer's manual.

In vitro and vivo S-nitrosylation assays. Recombinant proteins were *in vitro* S-nitrosylated with the stated concentration of the given nitric oxide donor in 500 μ l volumes for 20 min in darkness. Donors were removed using Micro Bio-Spin P6 columns (BioRad) and the resulting proteins were subjected to the biotinswitch technique²² by western blot assay using anti-biotin antibody (New England Biolab). For *in vivo* assay, *Arabidopsis* plants were inoculated with *Pst* DC3000 (*avrB*) at 10⁷ c.f.u. ml⁻¹ and an anti-Myc9 antibody (Sigma) was used.

Mass spectrometry. All chemicals were purchased from Sigma-Aldrich unless otherwise stated. Acetonitrile and water for liquid chromatography tandem mass spectrometry and sample preparation were of high-pressure liquid chromatography quality (Fisher). Formic acid was Suprapure 98–100% (Merck) and trifluoroacetic acid was 99% purity sequencing grade. Sequencing-grade modified porcine trypsin was purchased from Promega and GluC from Worthington (Lorne Lab). All high-pressure liquid chromatography mass spectrometry (LC-MS) connector fittings were from Upchurch Scientific or Valco (Hichrom and RESTEK).

As S-nitrosothiol formation at Cys 890 of AtRBOHD was found to be relatively labile *in vitro*, we used a well established method³⁶ that utilizes iodoacetamide to form a carbamidomethyl ion at sites of S-nitrosothiol formation, which are not blocked by treatment with methyl methanethiosulphonate (MMTS). This analysis revealed striking S-nitrosylation at Cys 890 of AtRBOHD but the complete absence of S-nitrosothiol formation at Cys 825 following treatment with a nitric oxide donor.

The GST-AtRBOHD C terminal was expressed in *E. coli* BL21 and purified using glutathione Sepharose 4B (GE Healthcare) in native condition. The purified protein solution was desalted using a Zeba Desalt spin column (Thermo Scientific). The desalted protein was treated with or without Cys-NO (final concentration, 0.5 mM) in HENS buffer for 20 min at room temperature (25 °C). Cys-NO was then removed using a Zeba desalt column and the free cysteines were blocked by MMTS in HENS buffer with 2.5% SDS for 20 min at 50 °C. The treated protein was precipitated by acetone and resuspended in HENS buffer with 1% SDS. Sodium ascorbate (10 mM) and iodoacetamide (50 mM) were added to remove S-NO bonds and for protein acylation. The proteins were separated by SDS-PAGE and the protein gel was excised, cleaned and digested with GluC and Trypsin at 37 °C for 16 h. The digested peptides were blocked first with MMTS and analysed by LC-MS.

Capillary gas high-pressure liquid chromatography tandem mass spectrometry (MSMS) analysis was performed on an online system consisting of a micropump (1200 binary HPLC system, Agilent) coupled to a hybrid LTQ-Orbitrap XL instrument (Thermo-Fisher). The LTQ was controlled through XCALIBUR 2.0.7. Samples were reconstituted in 10 μ l loading buffer before injection and analysed on a 1-h gradient for data-dependent analysis.

MSMS data were searched using MASCOT, versions 2.2 and 2.3 (Matrix Science Ltd), against a small database comprising the most common contaminant and various constructs. Variable methionine oxidation and cysteine methylthiol and carbamidomethylation were considered in all analyses. The precursor mass tolerance was set to 7 p.p.m. and the MSMS tolerance was set to 0.4 AMU. Fragmentation patterns were confirmed using ProteinProspector (<http://prospector.ucsf.edu>). Label-free quantitation was performed using PROGENESIS (Nonlinear Dynamics). For label-free quantitation, the number of 'features' (that is, signal at a specific retention time and *m/z*) was reduced to only MSMS peaks with a charge of 2, 3 or 4+ and only the five most intense MSMS spectra per feature were kept. Sets of multicharged ions (2+, 3+, 4+) were extracted from each LC-MS run.

FAD-binding activity assay. FAD-binding activity was measured as described previously³⁷. Briefly, proteins purified using a MagneGST Protein Purification System (Promega) were incubated with either GSNO or GSH for 20 min in the dark, and the excess GSNO or GSH was removed using a Bio-Spin6 column (Bio-Rad). The resulting compounds were further incubated with FAD (250 μ M) for 30 min in the dark. Unbound FAD was removed using a Bio-Spin6 column, and

the FAD content was determined by boiling the resulting protein samples for 5 min in the dark, followed by centrifugation at 14,000g for 10 min to remove coagulated protein. The absorbance of the released FAD was measured at 450 nm.

Computational modelling of NOX2 and AtRBOHD. The C-terminal 287 amino acids of NOX2 were used for identifying structural homologues with Phyre³⁸. This identified PDB crystal structure 1FDR, encoding an *E. coli* flavodoxin reductase, as a potential homologue. Structural alignments were optimized using DeepView and submitted for threading over 1FDR at the SwissModel server³⁹. The resulting computational model contained several loop regions unique to NOX2. These regions were analysed for secondary structure using Jpred⁴⁰ and the DeepView loop library. The computational model of the AtRBOHD C terminus was built by threading over both 1FDR and NOX2.

The model indicates that Cys 890 is positioned closely behind the conserved Phe 921 and Phe 570 residues in AtRBOHD and NOX2, respectively. Similar to the homologous Tyr247 in ferredoxin reductase²⁴, these residues are expected to have a significant role in FAD binding. Accordingly, mutation of Phe 570 has been reported to impair NOX2 function²⁵. The model further predicts that S-nitrosylation of AtRBOHD at Cys 890 may disrupt the side-chain position of Phe 921, impeding FAD binding.

31. Foissner, I., Wendehenne, D., Langebartels, C. & Durner, J. *In vivo* imaging of an elicitor-induced nitric oxide burst in tobacco. *Plant J.* **23**, 817–824 (2000).
32. Whalen, M. C., Innes, R. W., Bent, A. F. & Staskawicz, B. J. Identification of *Pseudomonas syringae* pathogens of *Arabidopsis* and a bacterial locus determining avirulence on both *Arabidopsis* and soybean. *Plant Cell* **3**, 49–59 (1991).
33. Dellagi, A., Brisset, M.-N., Jean-Pierre Paulin, J.-P. & Expert, D. Dual role of desferrioxamine in *Erwinia amylovora* pathogenicity. *Mol. Plant Microbe Interact.* **11**, 734–742 (1998).
34. Liu, Q., Li, M., Leibham, D., Cortez, D. & Elledge, S. The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol.* **8**, 1300–1309 (1998).
35. Sagi, M. & Fluhr, R. Superoxide production by plant homologues of the gp91^{phox} NADPH oxidase. Modulation of activity by calcium and by tobacco mosaic virus infection. *Plant Physiol.* **126**, 1281–1290 (2001).
36. Chen, Y. Y., Huang, Y. F., Khoo, K. H. & Meng, T. C. Mass spectrometry-based analyses for identifying and characterizing S-nitrosylation of protein tyrosine phosphatases. *Methods* **42**, 243–249 (2007).
37. Shen, A. L. & Kasper, C. B. Differential contribution of NADPH-cytochrome P450 oxidoreductase FAD binding site residues to flavin binding and catalysis. *J. Biol. Chem.* **275**, 41087–41091 (2000).
38. Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363–371 (2009).
39. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).
40. Cole, C., Barber, J. D. & Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**, W197–W201 (2008).

CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing

Sanjeev Shukla¹, Ersen Kavak^{2,3}, Melissa Gregory¹, Masahiko Imashimizu⁴, Bojan Shutinoski¹, Mikhail Kashlev⁴, Philipp Oberdoerffer¹, Rickard Sandberg^{2,3} & Shalini Oberdoerffer¹

Alternative splicing of pre-messenger RNA is a key feature of transcriptome expansion in eukaryotic cells, yet its regulation is poorly understood. Spliceosome assembly occurs co-transcriptionally, raising the possibility that DNA structure may directly influence alternative splicing. Supporting such an association, recent reports have identified distinct histone methylation patterns, elevated nucleosome occupancy and enriched DNA methylation at exons relative to introns. Moreover, the rate of transcription elongation has been linked to alternative splicing. Here we provide the first evidence that a DNA-binding protein, CCCTC-binding factor (CTCF), can promote inclusion of weak upstream exons by mediating local RNA polymerase II pausing both in a mammalian model system for alternative splicing, CD45, and genome-wide. We further show that CTCF binding to CD45 exon 5 is inhibited by DNA methylation, leading to reciprocal effects on exon 5 inclusion. These findings provide a mechanistic basis for developmental regulation of splicing outcome through heritable epigenetic marks.

It is estimated that greater than 90% of human genes undergo alternative splicing of pre-mRNA^{1,2} and aberrant splicing has been implicated in a number of human diseases³. Alternative splicing decisions are determined by the ability of weak splice sites to effectively compete with strong splice sites for detection by the spliceosome⁴. The balance between splice site selection is principally influenced by two variables⁵: (1) the availability of splicing factors that detect enhancer or silencer sequences encoded within nascent RNA^{4,6} and (2) the rate of RNA polymerase II (pol II) transcription elongation, wherein a slow rate favours co-transcriptional spliceosome assembly at weak splice sites^{7,8}.

A surprising result of genome-wide chromatin-immunoprecipitation-sequencing (ChIP-seq) studies is the non-random distribution of several epigenetic marks in exons relative to introns. In particular, exons show elevated nucleosome density, DNA methylation of cytosine, and overrepresentation of certain histone modifications, relative to introns^{9–15}. Differential ‘marking’ of exons on DNA highlighted a possible connection between DNA structure and co-transcriptional RNA processing. Accordingly, several recent studies suggest that exonic histone modification may affect variable inclusion of alternative exons^{16,17}. Collectively, these studies raise the intriguing possibility that epigenetic modifications are maintained on DNA to aid the spliceosome in the process of exon definition^{9,13}, and that differential chromatin assembly may represent a critical aspect of alternative splicing regulation.

Processing of CD45 pre-mRNA (also known as *PTPRC*) is a well established model system to study the regulatory mechanisms of alternative splicing. CD45 is a trans-membrane protein tyrosine phosphatase that initiates signalling through antigen receptors by dephosphorylating the inhibitory tyrosine on Src family kinases¹⁸. Variable exclusion of exons 4–6 (A–C) of CD45 transcripts is tightly correlated with stages of lymphocyte development and expressed splice variants can be distinguished using isoform-specific antibodies and flow cytometry¹⁹ (Supplementary Fig. 1). In general, the larger, exon 4-containing isoforms (CD45RA) are expressed early in peripheral lymphocyte

development, whereas the shortest isoform (CD45RO), which lacks all three variable exons, is expressed in terminally differentiated lymphocytes¹⁹. We recently identified heterogeneous ribonucleoprotein L-like (hnRNPLL) as a tissue-specific master regulator of the CD45RA to CD45RO transition in peripheral lymphocytes²⁰. hnRNPLL binds to exons 4 and 6 of CD45 mRNA and blocks the inclusion of both exons in the mature message²¹. In contrast, hnRNPLL expression does not influence exclusion of exon 5 (refs 20, 22; Supplementary Fig. 2a). *In vitro* studies aimed at uncovering regulators of exon 5 exclusion have identified several ubiquitously expressed splicing factors^{22,23}. However, peripheral lymphocytes retain exon 5 until the terminal stages of development²⁴ (Supplementary Fig. 2b), portending a yet uncovered layer of regulation.

CTCF regulates exon 5 inclusion in CD45 mRNA

Considering the growing evidence for DNA-mediated regulation of spliceosome assembly, we explored the hypothesis that exon 5 inclusion is mediated by the epigenetic structure of the gene encoding CD45. By analysing published ChIP-seq data within the UCSC genome browser^{25,26}, we identified a strong CTCF peak overlapping with exon 5 across cell types. CTCF is an 11 zinc-finger DNA-binding protein with multiple nuclear functions, largely grouped into two categories: insulating inactive regions of the genome from active regions and promoting long-range interactions between distal regions of the genome^{27,28}. Whereas intergenic CTCF is an effective barrier to transcription²⁷, we found that CTCF binding at exon 5 is maintained in cells that actively transcribe abundant CD45 (ref. 26) and that exon 5 binding is conserved in murine splenocytes (Fig. 1a and Supplementary Fig. 2c), indicating an important, position-dependent ‘non-insulator’ function. We thus explored whether and how CTCF binding at CD45 exon 5 DNA influences processing of CD45 transcripts.

To dissect the impact of CTCF on exon 5 splicing in a cell-based system, we screened several human Burkitt lymphoma B cell lines for differences in expression of the exon 5 containing CD45RB isoform.

¹Center for Cancer Research, Mouse Cancer Genetics Program, National Cancer Institute at Frederick, Frederick, Maryland 21702, USA. ²Department of Cell and Molecular Biology, Karolinska Institutet, SE-171 77 Stockholm, Sweden. ³Ludwig Institute for Cancer Research, SE-171 77, Stockholm, Sweden. ⁴Center for Cancer Research, Gene Regulation and Chromosome Biology Laboratory, National Cancer Institute at Frederick, Frederick, Maryland 21702, USA.

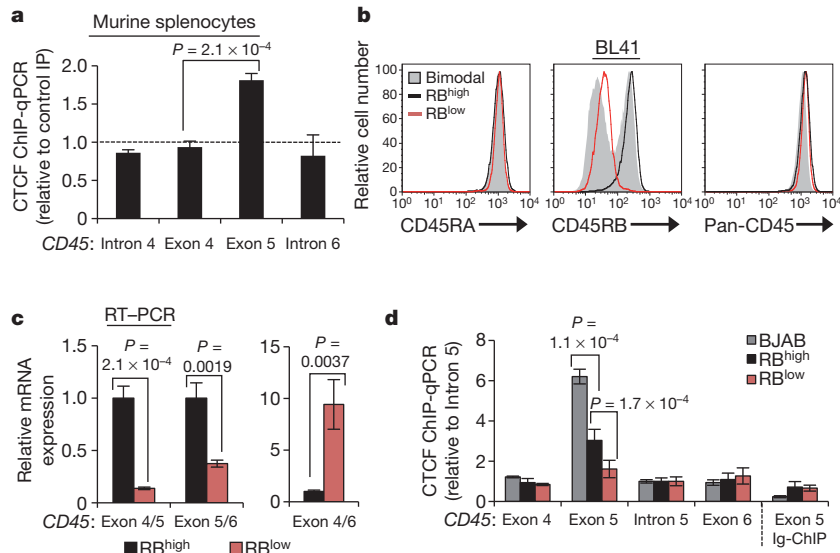


Figure 1 | Binding of CTCF to exon 5 of *CD45* DNA is associated with inclusion of exon 5 in *CD45* transcripts. **a**, CTCF ChIP in murine splenocytes and quantitative PCR (qPCR) relative to rabbit Ig control ChIP ($n = 2$). **b**, Cell-surface staining for CD45RA (exon 4-containing) and CD45RB (exon 5-containing) isoforms and total CD45 (pan) in parental BL41 B cells (RB^{high}), cell-culture-derived CD45RB bimodal cells, and CD45RB low (RB^{low}) cells

Whereas lymphocyte cell lines generally express high levels of CD45RB (Supplementary Fig. 2d, e), culturing BL41 cells in non-heat-inactivated fetal bovine serum (FBS) resulted in bimodal CD45RB expression. CD45RB low cells (RB^{low}) were sorted from the bimodal population and stably maintained (Fig. 1b). Parental BL41 cells (RB^{high}) and sorted RB^{low} cells express equivalent exon 4-containing CD45RA isoforms and total CD45 (Fig. 1b), indicating specific exclusion of exon 5 in the RB^{low} population. Quantitative RT-PCR with exon junction spanning primers validated exon 5 skipping: RB^{low} cells showed reduced exon 4/5 and exon 5/6 junctions, but enhanced exon 4/6 junctions relative to RB^{high} cells (Fig. 1c). Notably, several histone modifications that have been previously linked to alternative splicing (H3K36me3, H3K27me3, H3K4me3)^{16,29} are equivalently detected at exon 5 in RB^{high} and RB^{low} cells (Supplementary Fig. 3a, b). CTCF-ChIP in the newly identified RB^{high} and RB^{low} BL41 cells, and CD45RB-high BJAB cells revealed a strong positive correlation between exon 5 inclusion in *CD45* mRNA and CTCF binding at *CD45* exon 5 DNA (Fig. 1b–d and Supplementary Fig. 2e), particularly in BJAB cells, which also express elevated CTCF protein (Supplementary Fig. 3c). In agreement with the observation that exon 5 splicing is independent of hnRNPLL, modulation of hnRNPLL expression did not influence CTCF binding to *CD45* exon 5 (Supplementary Fig. 3d).

To assess whether the association between CTCF binding and exon 5 inclusion reflects a direct role for CTCF in *CD45* alternative splicing, we used RNA interference to deplete CTCF from our B cell lines (Supplementary Fig. 4a). Decreasing CTCF levels in bimodal BL41 cells led to a marked loss of CD45RB expression without reducing overall CD45 levels (Fig. 2a and Supplementary Fig. 4b). Similarly, CTCF depletion in RB^{low} cells and BJAB cells led to a substantial loss of CD45RB staining with little effect on overall CD45 levels (Fig. 2a). Quantitative RT-PCR of *CD45* mRNA in CTCF-depleted RB^{low} and BJAB cells validated reduced exon 5 expression and increased exon 4/6 junctions (Fig. 2b, c, respectively; Supplementary Fig. 4c, additional transductions), confirming that CTCF mediates exon 5 inclusion.

CTCF promotes pol II pausing at *CD45* exon 5

We next investigated the mechanism by which CTCF binding to *CD45* DNA influences mRNA splicing outcomes. Given that genome-wide ChIP-seq studies have revealed overlapping intragenic CTCF and pol

II peaks³⁰, we examined whether CTCF promotes inclusion of exon 5 through interference with pol II elongation. ChIP confirmed significant enrichment of pol II at *CD45* exon 5 DNA, but not at adjacent regions in RB^{high} cells as compared to RB^{low} cells (Fig. 3a). Using antibodies specific to pol II phosphorylated on the carboxy-terminal domain (CTD), we further showed that elevated pol II at *CD45* exon 5 in RB^{high} cells is associated with the elongating form phosphorylated on serine 2 of CTD YSPTSPS heptad repeats³¹ (Supplementary Fig. 5a, b). Notably, CTCF depletion from RB^{high} cells (Supplementary Fig. 5c, d) reduced both CTCF binding (Fig. 3b) and pol II levels at *CD45* exon 5 (Fig. 3c).

II peaks³⁰, we examined whether CTCF promotes inclusion of exon 5 through interference with pol II elongation. ChIP confirmed significant enrichment of pol II at *CD45* exon 5 DNA, but not at adjacent regions in RB^{high} cells as compared to RB^{low} cells (Fig. 3a). Using antibodies specific to pol II phosphorylated on the carboxy-terminal domain (CTD), we further showed that elevated pol II at *CD45* exon 5 in RB^{high} cells is associated with the elongating form phosphorylated on serine 2 of CTD YSPTSPS heptad repeats³¹ (Supplementary Fig. 5a, b). Notably, CTCF depletion from RB^{high} cells (Supplementary Fig. 5c, d) reduced both CTCF binding (Fig. 3b) and pol II levels at *CD45* exon 5 (Fig. 3c).

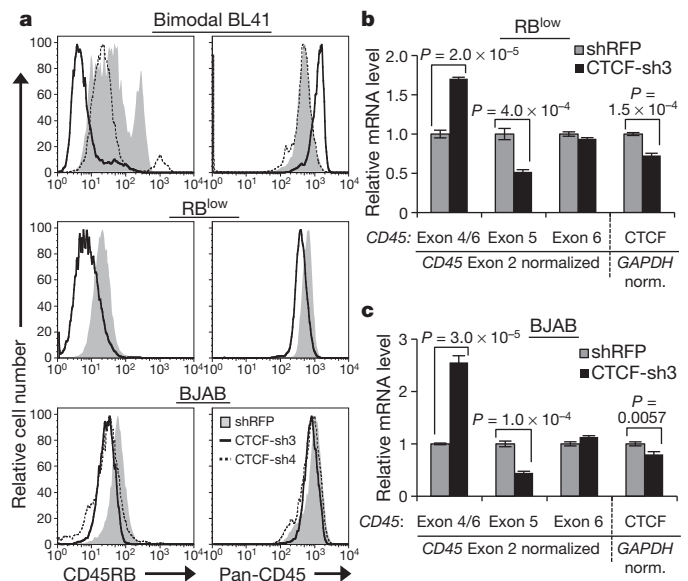


Figure 2 | CTCF depletion leads to reduced exon 5 inclusion in *CD45* transcripts. **a**, Cell-surface CD45RB isoform and total CD45 expression in cells transduced with short hairpin RNA (shRNA) against CTCF (CTCF-sh3 and/or sh-4) or control shRNA against red fluorescent protein (RFP). **b**, **c**, qRT-PCR in CTCF-depleted RB^{low} (**b**) and BJAB cells (**c**) from **a** to detect *CD45* (left) and *CTCF* (right) mRNA levels ($n = 3$). Graphs show mean values \pm s.d. P , two-tailed Student's t test.

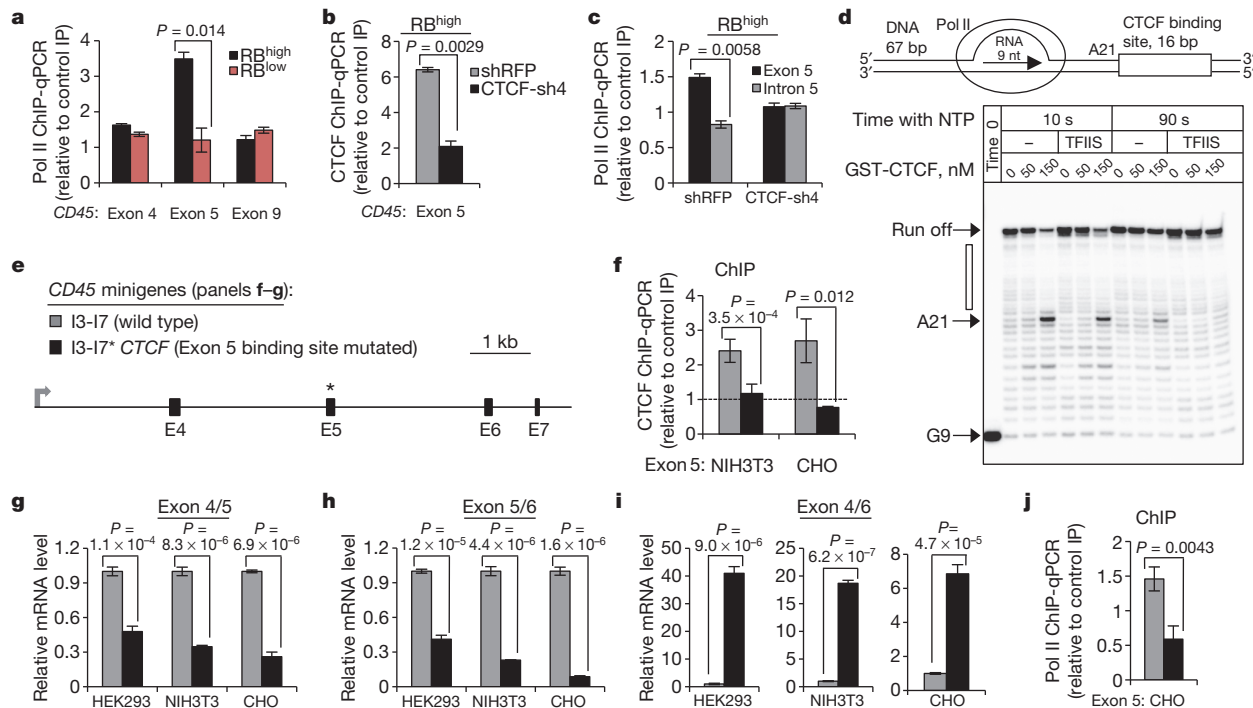


Figure 3 | CTCF binding at *CD45* exon 5 DNA facilitates exon 5 inclusion in *CD45* transcripts through local pol II pausing. **a**, RNA pol II ChIP and qPCR relative to mouse Ig control IP ($n = 3$). **b**, CTCF ChIP in RB^{high} cells transduced with shRNA against *CTCF* versus shRFP-transduced cells and qPCR relative to rabbit Ig control IP ($n = 2$). **c**, RNA pol II ChIP of RB^{high} cells from **b** and qPCR relative to mouse Ig control IP ($n = 2$). **d**, *In vitro* transcription with a DNA oligo incorporating a CTCF binding site at position 26 relative to elongation complex assembly. Recombinant CTCF and TFIIS protein were introduced as indicated, with variable effects on pausing at adenine 21 (A21).

The above data definitely link CTCF binding, pol II pausing and exon 5 inclusion, but do not exclude additional, context-dependent secondary effects. To query whether CTCF binding to an actively transcribed template is sufficient to promote pol II pausing, we assembled a pol II ternary elongation complex from synthetic DNA and RNA oligonucleotides and highly purified yeast pol II³². A CTCF binding site was incorporated into the template DNA at position 26 relative to the hybridization location of a 9-nucleotide RNA primer (Fig. 3d). CTCF binding to the target sequence was confirmed by electrophoretic mobility shift assay (EMSA) (Supplementary Fig. 6a). Incubation with pol II and increasing amounts of recombinant CTCF resulted in pausing immediately upstream of the CTCF binding site (Fig. 3d). Extended incubation or introduction of the elongation factor TFIIS substantially reduced pausing and led to near complete escape of paused pol II (Fig. 3d and Supplementary Fig. 6b). Thus, CTCF can autonomously promote pol II pausing, but not complete arrest, on a naked DNA template. These data establish CTCF as a direct impediment to transcription that can act in the absence of a particular nucleosome structure or chromatin context. Furthermore, the ability of paused pol II to resume transcription efficiently in the presence of CTCF supports a physiological role for CTCF in favouring exon inclusion through transient, spatiotemporal pol II pausing.

Having demonstrated that CTCF can promote pol II pausing, we explored the relationship between CTCF, pol II and exon inclusion in a tractable, endogenous system. We generated a wild-type minigene extending from intron 3 through intron 7 of human *CD45* genomic DNA (I3-17), as well as a mutant analogue, in which the exon 5 CTCF binding site was disrupted through nucleotide substitution (I3-17*CTCF) (Fig. 3e and Supplementary Fig. 6c). The 11 zinc fingers of CTCF support multiple contacts to substrate DNA^{33,34} and a minimum of five substitutions within the core motif³⁵ were required to

significantly ablate CTCF binding (EMSA, Supplementary Fig. 6d). To avoid detection of endogenous *CD45*, which is confined to the haematopoietic lineage¹⁸, the minigenes were transfected into several fibroblast cell lines. In addition to human HEK293 cells, murine NIH3T3 and hamster CHO cells were used to specifically amplify human minigene *CD45* DNA in ChIP analyses. CTCF ChIP of transfected NIH3T3 and CHO cells confirmed robust binding to exon 5 of the I3-17 minigene, and complete disruption of binding to the mutated, I3-17*CTCF minigene (Fig. 3f). Quantitative RT-PCR indicated that both minigenes were comparably expressed and approximated endogenous *CD45* levels in immune cells (Supplementary Fig. 6e). Mutation of the CTCF binding site in exon 5 led to a marked decrease in 4/5 and 5/6 junctions, and increase in 4/6 junctions in all three cell types (Fig. 3g, h, i, respectively), resulting in an overall 50–100× decrease in exon 5 inclusion. Notably, ChIP confirmed increased pol II occupancy at exon 5 in the I3-17 minigene, but not in the mutated, I3-17*CTCF minigene (Fig. 3j). As the two minigenes are identical in every regard minus the five core nucleotides of the CTCF binding site, these data establish CTCF as a direct regulator of *CD45* exon 5 inclusion, which operates through promoting local pol II pausing.

DNA methylation inhibits exon 5 CTCF binding

Armed with the knowledge that CTCF binding to exon 5 DNA regulates inclusion, and given that exon 5 is variably excluded during lymphocyte maturation, we asked whether and how CTCF binding is modulated to influence splicing outcome. Whereas CTCF is ubiquitously expressed, binding to DNA is inhibited by methylation on CpG dinucleotides^{27,34}. Several recent studies have shown that DNA methylation is substantially enriched at exons relative to introns^{14,15,36}, suggesting a role in pre-mRNA processing, yet a causal relationship between these processes has not been demonstrated. Methylated

DNA immunoprecipitation (MedIP) in our B-cell lines suggested that CTCF binding at *CD45* exon 5 and associated exon inclusion are indeed regulated by DNA methylation: we detected a strong inverse correlation between CTCF and 5-methylcytosine at *CD45* exon 5, but not at adjacent exons (Figs 1d, 4a). To assess whether DNA methylation of *CD45* exon 5 and reciprocal loss of CTCF binding contribute to exon 5 exclusion during the transition from naïve to mature T lymphocytes, $CD3^+$ T cells were isolated from human peripheral blood and sorted into RB^{high} (naïve) and RB^{medium} (mature) populations¹⁹ (Fig. 4b). MedIP confirmed significant enrichment of *CD45* exon 5 methylation (Fig. 4c) and reduced CTCF binding (Fig. 4d) in RB^{medium} cells compared to RB^{high} peripheral T cells. Thus, CTCF binding and *CD45* exon 5 inclusion are inversely related to DNA methylation in several transformed cell lines and in primary T cells.

To determine whether dynamic methylation of *CD45* exon 5 DNA is a regulatory mechanism contributing to *CD45* alternative splicing, we modulated methylation through inhibition of the DNA maintenance methyltransferase, DNMT1. We reasoned that, if elevated exon 5 methylation and consequent reduced CTCF binding were the principal components distinguishing RB^{low} and RB^{high} cells, inhibition of methylation should cause RB^{low} cells to revert to an RB^{high} phenotype. Indeed, DNMT1 depletion in RB^{low} cells (Supplementary Fig. 7a, b) reduced 5-methylcytosine levels (Fig. 4e) and restored CTCF binding at *CD45* exon 5 (Fig. 4f), leading to enhanced exon 5 inclusion in *CD45* mRNA, as evidenced by increased exon 4/5 and 5/6 junctions (Supplementary Fig. 7c) and cell-surface CD45RB (Fig. 4g). Notably, increasing CTCF binding in RB^{low} cells through reduced exon 5 methylation also reinstated local pol II pausing (Fig. 4h). In addition to identifying dynamic DNA methylation as a possible regulatory mechanism governing *CD45* alternative splicing *in vivo*, these data establish CTCF as the first mechanistic link between DNA methylation and alternative pre-mRNA splicing.

Global effects of intragenic CTCF on splicing

Although studies of CTCF function have been largely restricted to intergenic activities, CTCF ChIP-seq studies found that approximately 40–45% of CTCF binding sites are located intragenically^{35,37,38}. Based on our observations with *CD45*, we propose that some portion of intragenic CTCF binding sites operate to influence pre-mRNA processing decisions. To globally address the impact of CTCF on alternative splicing, we performed CTCF ChIP-seq in BL41 and

BJAB cells to produce cell-type-specific CTCF binding maps, and high-throughput RNA-sequencing (RNA-seq) of total RNA from CTCF-depleted BL41 and BJAB cells and their relevant controls (CTCF-sh3, Fig. 2a). Mapping of overall CTCF binding sites in BL41 and BJAB cells indicated comparable distribution patterns to previous reports (Supplementary Table 2). The mixture of isoforms (MISO) model was applied to RNA-seq data (Supplementary Table 3) to identify exons with a high probability of differential expression in response to CTCF depletion, as assessed by the Bayes factor confidence index³⁹. Exons showing altered inclusion in response to CTCF depletion were further subdivided into three categories based on proximity to a local CTCF binding site: unbound by CTCF, or CTCF-bound within 1 kilobase downstream or upstream of the exon (Fig. 5a and Supplementary Table 4). CTCF is a global regulator of transcription²⁷ and depletion would be expected to result in some level of alternative splicing due to alterations in upstream pathways. Accordingly, MISO identified exons that were differentially included in mRNA in response to CTCF depletion, but were not locally bound by CTCF on the corresponding DNA. Importantly, in BL41 and BJAB cells, alternative exons not bound by CTCF were centred at zero across Bayes factor thresholds, indicating that secondary effects of CTCF depletion showed no preference towards exon inclusion or exclusion (Fig. 5b, c, respectively). Similarly, CTCF binding upstream of the differentially expressed exon did not show a statistically significant bias towards exon inclusion or exclusion (Fig. 5b, c and Supplementary Fig. 8a, b). However, we detected a strong correlation between CTCF depletion and exon exclusion if CTCF is bound downstream of the alternative exon in both BL41 and BJAB cells (Fig. 5b, c and Supplementary Fig. 8a, b). We additionally identified CTCF-bound exons that showed reduced inclusion in BL41 and BJAB cells, as well as unique examples, indicating a degree of cell-type specificity (Supplementary Figs 8c, 9a and Supplementary Table 4).

These genome-level data are consistent with our observations in the *CD45* model system, wherein CTCF binding downstream of the weak 3' splice site flanking exon 5 promoted inclusion of exon 5 in mature message, but had no effect on exon 6 (Figs 1c, d and 2b, c). As we had mechanistically linked CTCF-associated pol II pausing to *CD45* exon 5 inclusion, we examined pol II occupancy at the downstream CTCF sites that led to reduced exon inclusion upon CTCF depletion. Inspection of publicly available CTCF ChIP-seq data from $CD4^+$ T cells³⁷ indicated high conservation of these CTCF binding sites²⁶.

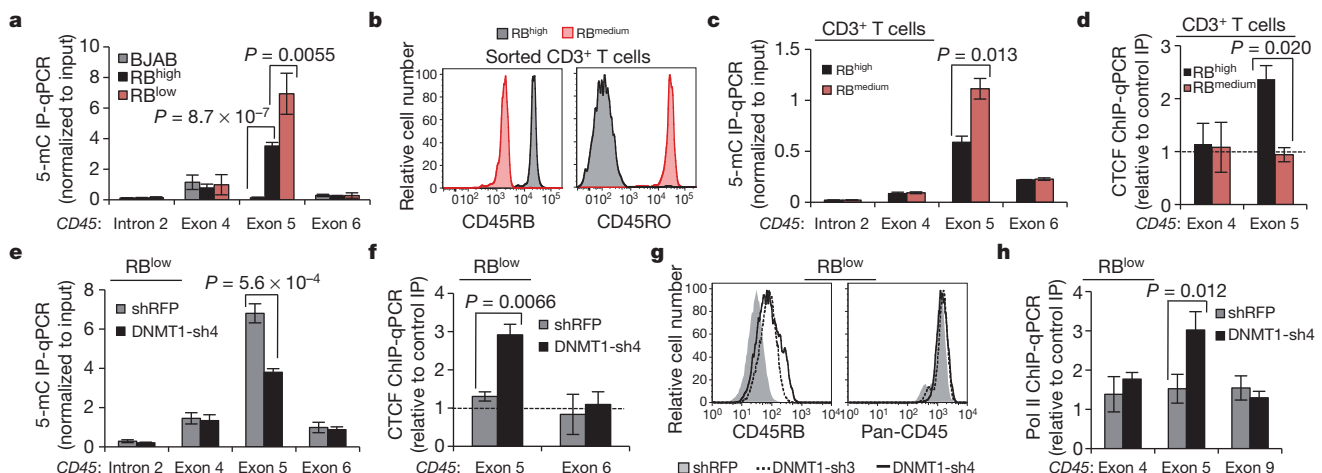


Figure 4 | 5-methylcytosine levels (5-mC) are inversely related to CTCF binding and exon 5 inclusion. **a**, Methylated DNA immunoprecipitation (MedIP) in B cell line genomic DNA and qPCR relative to input ($n = 5$). **b**, Representative *CD45* isoform expression in primary peripheral human $CD3^+$ T cells sorted on the basis of cell-surface *CD45RB* and *CD45RO*. **c**, MedIP and qPCR relative to input in sorted primary human $CD3^+$ T cells ($n = 6$, compiled from two donors). **d**, CTCF-ChIP and qPCR relative to rabbit

Ig control IP, in sorted primary $CD3^+$ T cells ($n = 2$). **e**, MedIP and qPCR relative to input in BL41 RB^{low} cells transduced with shRNA against *DNMT1* versus shRFP-transduced cells ($n = 3$). **f**, CTCF ChIP in cells from **e** and qPCR relative to rabbit Ig control IP ($n = 3$). **g**, Cell-surface *CD45RB* expression in cells from **e**. **h**, RNA pol II ChIP and qPCR in cells from **e** relative to mouse Ig control IP ($n = 3$). Unless indicated otherwise, graphs show mean values \pm s.d. *P*, two-tailed Student's *t* test.

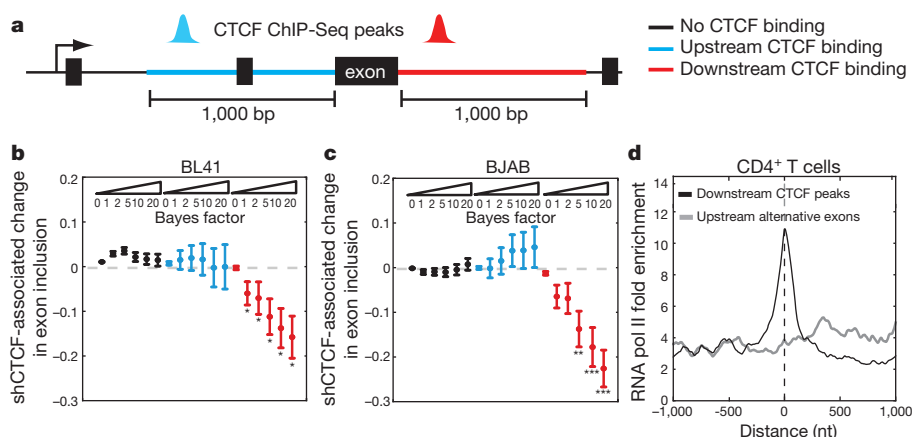


Figure 5 | Global identification of CTCF-dependent exons. **a**, Alternative exons were classified on the basis of the relative location of an exclusive CTCF peak within 1 kb of the exon. **b**, Difference in the mean exon inclusion level between bimodal BL41 cells transduced with shRNA against CTCF versus shRFP-transduced cells (from Fig. 2a) for exons with CTCF peak in upstream (blue) or downstream regions (red) but not in the exon body and for exons with no CTCF binding (black). The mean \pm s.e.m. for each class of exons is plotted

against increasing Bayes factor thresholds. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, Wilcoxon rank sum test for differences in exon inclusion at the different thresholds. **c**, Same as **b** for BJAB shCTCF compared with wild-type BJAB cells (from Fig. 2a). **d**, Normalized CD4⁺ T cell RNA pol II read signal centred on the alternative exon or the corresponding downstream CTCF peak summit.

Analysis of the corresponding pol II ChIP-seq data³⁷ revealed a stronger enrichment of pol II at downstream CTCF binding sites relative to upstream exons (Fig. 5d). Enrichment of pol II occupancy at CTCF binding sites compared to associated upstream alternative exons was confirmed for several genes in BJAB and BL41 cells (Supplementary Fig. 9b). Together with our CD45 data, we conclude that CTCF bound downstream of alternative exons promotes pol II pausing, providing the necessary temporal context for co-transcriptional spliceosome assembly at weak upstream splice sites.

Discussion

In recent years, the link between DNA structure and pre-mRNA processing has been gaining increasing attention. Reports of increased nucleosome occupancy and DNA methylation as well as distinct histone methylation patterns at exons relative to introns have fuelled the hypothesis that exons are differentially marked to aid the spliceosome in the process of exon definition^{16,17}. It has further been shown that pol II occupancy increases in the vicinity of exons, although whether a function of DNA sequence, chromatin structure or the presence of DNA-binding proteins has not been defined. Recently, several studies have linked modification of distinct histone methylation patterns to alternative splicing^{16,17}. However, exonic histone methylation was shown to be equivalent in other models of robust exon inclusion versus exclusion⁴⁰, suggesting that, although histone methylation patterns may prime splicing decisions, they probably do so in concert with other factors. Consistent with the latter, we observed comparable histone methylation at exon 5 whether or not exon 5 was included in the CD45 message (Supplementary Fig. 3a, b). Rather, we show that mutually exclusive DNA methylation and CTCF binding regulate exon 5 inclusion through influencing pol II elongation dynamics (Supplementary Fig. 10). Given that mapping of CTCF binding sites shows roughly 40–70% conservation between tissues²⁷, it is tempting to speculate that altered DNA methylation patterns during development can lead to variations in intragenic CTCF binding that thereby contribute to tissue-specific alternative splicing patterns. This may be especially relevant in pathological conditions, such as cancer, where widespread changes in DNA methylation, altered CTCF binding, and aberrant alternative pre-mRNA splicing have been reported^{15,41–43}. We predict that our identification of CTCF as a DNA-binding regulator of alternative pre-mRNA splicing represents the tip of the iceberg, and that a long list of location-specific DNA-binding ‘splicing factors’ will follow.

METHODS SUMMARY

Experiments were performed with BJAB and BL41 cells or primary lymphocytes. CD45 isoform analysis was achieved with isoform-specific antibodies or pan-antibody directed against a common region of CD45. Transductions were executed with vesicular stomatitis virus G (VSV-G)-pseudotyped lentivirus, and selected for puromycin resistance. Quantitative RT-PCR was performed on cDNA from total RNA. Protein lysates were prepared with RIPA buffer. ChIP and MedIP were conducted with formaldehyde cross-linked, sonicated material. *In vitro* transcription elongation was performed with yeast RNA pol II, yeast TFIIIS and human CTCF. Minigenes were cloned into the pCI-neo (Promega) construct and transfected with Lipofectamine 2000 (Invitrogen). ChIP-Seq and RNA-Seq were executed with the Illumina platform. For ChIP-Seq, Illumina FastQ files were mapped to the human genome (hg19). Peak calling was run using Rabbit Ig control sequencing data as background. For RNA-Seq, exon inclusion levels were determined using the MISO program³⁹.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 March; accepted 12 August 2011.

Published online 2 October 2011.

- Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **40**, 1413–1415 (2008).
- Tazi, J., Bakkour, N. & Stamm, S. Alternative splicing and disease. *Biochim. Biophys. Acta* **1792**, 14–26 (2009).
- Matlin, A. J., Clark, F. & Smith, C. W. Understanding alternative splicing: towards a cellular code. *Nature Rev. Mol. Cell Biol.* **6**, 386–398 (2005).
- Han, J., Xiong, J., Wang, D. & Fu, X. D. Pre-mRNA splicing: where and when in the nucleus. *Trends Cell Biol.* **21**, 336–343 (2011).
- Singh, R. & Valcarcel, J. Building specificity with nonspecific RNA-binding proteins. *Nature Struct. Mol. Biol.* **12**, 645–653 (2005).
- Kornblihtt, A. R. Coupling transcription and alternative splicing. *Adv. Exp. Med. Biol.* **623**, 175–189 (2007).
- de la Mata, M. *et al.* A slow RNA polymerase II affects alternative splicing *in vivo*. *Mol. Cell* **12**, 525–532 (2003).
- Tilgner, H. *et al.* Nucleosome positioning as a determinant of exon recognition. *Nature Struct. Mol. Biol.* **16**, 996–1001 (2009).
- Spies, N., Nielsen, C. B., Padgett, R. A. & Burge, C. B. Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* **36**, 245–254 (2009).
- Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. & Komorowski, J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* **19**, 1732–1741 (2009).
- Kolasinska-Zwier, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genet.* **41**, 376–381 (2009).
- Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon-intron structure. *Nature Struct. Mol. Biol.* **16**, 990–995 (2009).
- Chodavaram, R. K. *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature* **466**, 388–392 (2010).

15. Hodges, E. *et al.* High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res.* **19**, 1593–1605 (2009).
16. Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. & Misteli, T. Epigenetics in alternative pre-mRNA splicing. *Cell* **144**, 16–26 (2011).
17. Alló, M. *et al.* Chromatin and alternative splicing. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 103–111 (2010).
18. Trowbridge, I. S. & Thomas, M. L. CD45: an emerging role as a protein tyrosine phosphatase required for lymphocyte activation and development. *Annu. Rev. Immunol.* **12**, 85–116 (1994).
19. Hermiston, M. L., Xu, Z. & Weiss, A. CD45: a critical regulator of signaling thresholds in immune cells. *Annu. Rev. Immunol.* **21**, 107–137 (2003).
20. Oberdoerffer, S. *et al.* Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL. *Science* **321**, 686–691 (2008).
21. Topp, J. D., Jackson, J., Melton, A. A. & Lynch, K. W. A cell-based screen for splicing regulators identifies hnRNP LL as a distinct signal-induced repressor of CD45 variable exon 4. *RNA* **14**, 2038–2049 (2008).
22. Motta-Mena, L. B., Heyd, F. & Lynch, K. W. Context-dependent regulatory mechanism of the splicing factor hnRNP L. *Mol. Cell* **37**, 223–234 (2010).
23. Tong, A., Nguyen, J. & Lynch, K. W. Differential expression of CD45 isoforms is controlled by the combined activity of basal and inducible splicing-regulatory elements in each of the variable exons. *J. Biol. Chem.* **280**, 38297–38304 (2005).
24. Horgan, K. J. *et al.* CD45RB expression defines two interconvertible subsets of human CD4⁺ T cells with memory function. *Eur. J. Immunol.* **24**, 1240–1243 (1994).
25. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
26. UCSC Genome Browser, GRC37/hg19, ENCODE Histone Modification Tracks.
27. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
28. Ohlsson, R., Bartkuhn, M. & Renkawitz, R. CTCF shapes chromatin by multiple mechanisms: the impact of 20 years of CTCF research on understanding the workings of chromatin. *Chromosoma* **119**, 351–360 (2010).
29. Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science* **327**, 996–1000 (2010).
30. Wada, Y. *et al.* A wave of nascent transcription on activated human genes. *Proc. Natl Acad. Sci. USA* **106**, 18357–18361 (2009).
31. Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36**, 541–546 (2009).
32. Komissarova, N., Kireeva, M. L., Becker, J., Sidorenkov, I. & Kashlev, M. Engineering of elongation complexes of bacterial and yeast RNA polymerases. *Methods Enzymol.* **371**, 233–251 (2003).
33. Filippova, G. N. *et al.* An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell. Biol.* **16**, 2802–2813 (1996).
34. Renda, M. *et al.* Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J. Biol. Chem.* **282**, 33336–33345 (2007).
35. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–1245 (2007).
36. Lyko, F. *et al.* The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* **8**, e1000506 (2010).
37. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
38. Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of *in vivo* protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.* **36**, 5221–5231 (2008).
39. Katz, Y., Wang, E. T., Airolidi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods* **7**, 1009–1015 (2010).
40. Huff, J. T., Plocik, A. M., Guthrie, C. & Yamamoto, K. R. Reciprocal intronic and exonic histone modification regions in humans. *Nature Struct. Mol. Biol.* **17**, 1495–1499 (2010).
41. Berdasco, M. & Esteller, M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev. Cell* **19**, 698–711 (2010).
42. David, C. J. & Manley, J. L. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* **24**, 2343–2364 (2010).
43. Jelinic, P. & Shaw, P. Loss of imprinting and cancer. *J. Pathol.* **211**, 261–268 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Rao, C. Burge and K. Lynch for critical reading of this manuscript. We also thank A. Rao for reagents and K. Nyswaner and M. Prigge for technical assistance. This work is supported by the Intramural Research Program of NIH, the National Cancer Institute, The Center for Cancer Research (S.O., P.O., M.K.), and the Swedish Research Council Foundation and the Foundation for Strategic Research (R.S.).

Author Contributions S.S. performed ChIP, MedIP and EMSA. M.G. and S.S. performed lentiviral transductions, transfections, flow cytometry, and qPCR. E.K. analysed ChIP and RNA-seq data. M.I. performed *in vitro* transcription. S.S. and B.S. cloned the minigenes. All authors designed experiments and M.K., P.O., R.S. and S.O. supervised the project. S.O. and R.S. wrote the text. P.O. edited the text.

Author Information All data sets in this publication are available in the NCBI Gene Expression Omnibus accession number GSE31278. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.O. (shalini.oberdoerffer@nih.gov).

METHODS

Cell culture. BJAB and BL41 cells were maintained at 37 °C, 5% CO₂ in RPMI (Invitrogen) supplemented with 10% FBS (Hyclone), and 1% L-glutamine. BJAB and parental RB^{high} BL41 cells were cultured in heat-inactivated FBS, whereas RB^{low} BL41 cells were initially kept in native FBS, but were ultimately transitioned into inactivated serum. JSL1 cells were maintained at 37 °C, 5% CO₂ in RPMI (Invitrogen) supplemented with 5% FBS (Hyclone), and 1% L-glutamine. Primary human peripheral blood lymphocytes were purified by spinning through Ficoll Paque (GE Healthcare). Isolated cells were washed twice with PBS and CD3⁺ T cells were isolated with CD3⁺ microbeads (Miltenyi Biotech). Primary murine splenocytes were isolated from whole spleen of BL/6 mice. Single-cell suspensions were lysed with ACK lysis buffer (0.15 M NH₄Cl, 10 mM KHCO₃, 0.1 mM Na₂EDTA) to remove red blood cells before ChIP assay. JSL1 cells were stimulated at a concentration of 3×10^5 cells per ml. Phorbol 12-myristate 13-acetate (PMA) was added at a final concentration of 20 nM. Flow cytometry was performed 2 days post-stimulation.

Virus production. Constructs encoding shRNA directed against *CTCF* and *DNMT1* were obtained from Open Biosystems and were transfected (Lipofectamine 2000, Invitrogen) along with VSV-G and *gag/pol* (courtesy of The RNAi Consortium of the Broad Institute) into 293T cells for viral production. Viral supernatants were concentrated 50× and aliquoted for storage.

Cell line infection. BJAB and BL41 cells were plated in 96-well round-bottom plates at 100,000 cells per well. Five microlitre of virus and 8 µg ml⁻¹ polybrene were added per well and the plate was spun at 760g for 90 min. The supernatants were removed and fresh media was added. Puromycin was added at a final concentration of 5 µg ml⁻¹ on day 2. Depletion of CTCF from cells resulted in significant cell death after 1 week in culture and depletion of DNMT1 resulted in silencing after 10 days in culture. Cells for downstream analysis were collected 5 (shCTCF) or 7 (shDNMT1) days post-infection. To scale up infections for ChIP and western blotting, infections were performed in individual wells of 96-well plates and pooled before harvesting for RNA, ChIP and western blot. Three plates were pooled for shCTCF experiments and 3.5 plates were pooled for shDNMT1 experiments. Three individual RNA and ChIP samples were taken from each of the bulk cultures.

Target sequences of shRNAs. DNMT1-sh3, 5'-CGAGAAGAATATCGAAC TCTT-3'; DNMT1-sh4, 5'-CGACTACATCAAAGGCAGCAA-3'; CTCF-sh3, 5'-CCTCCTGAGGAATCACCTTAA-3'; CTCF-sh4, 5'-GCGGAAAGTGAA CCCATGATA-3'; shRFP, 5'-GAATTAAGAGAGGCTCAGTTA-3'; LL-sh4, 5'-CGACAGGCTCTAGTGAATTT-3'.

Flow cytometry. The following antibodies were used for flow cytometry: CD45RO clone UCHL1 (eBioscience, 12-0457-42, batch no. E034572), CD45RA clone MEM-56 (ExBio, 1P-223, batch no. 11827), CD45RB clone MT4 (BD Pharmingen, 555904, batch no. 89956) and pan-CD45 clone HI30 (BD Pharmingen, 555483, batch no. 555483). Staining of CD45 isoforms was performed in separate tubes, to avoid competition for antibody binding. Flow cytometry was performed on either a BD FACScalibur or BD LSR II cytometer.

Quantitative RT-PCR. RNA was isolated with the Qiagen RNeasy Mini Kit and reverse transcription was performed with SuperScript II (Invitrogen) according to the manufacturer's instructions. PCR measurements were performed in triplicate in the presence of SYBR green reagent (Roche) and amplification was performed on a 480 Light Cycler (Roche). The average cycle thresholds for the technical triplicates were calculated to yield one value per primer set for each biological replicate. Normalization was performed to *GAPDH*, *RPS16* or surrounding exon level values using the formula $2^{(C_{\text{normalization}} - C_{\text{experimental}})}$ to determine relative expression. Averages and standard deviations of the normalized biological replicate values were plotted in the figures and used in *t*-test calculations. Figure legends indicate the number of biological replicates (individual RNA preparations) used in each experiment.

Western blots. Cells were lysed in RIPA buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% NP40, 0.1% SDS, 0.5% sodium deoxycholate, and 1× Halt protease inhibitor cocktail (Thermo Scientific)). Proteins (35 µg) were loaded per lane on a 4–20% gradient SDS-PAGE gel. Western blot was performed with anti-CTCF clone D31H2 (Cell signaling 3418S, batch no. 1), DNMT1 antibody (Abcam ab13537, batch no. GR16960-1), or anti-p65 RelA (BD Bioscience 610869, batch no. 50886) antibodies. Anti-RelA immunoblotting served as a loading control for protein levels.

Chromatin immunoprecipitation (ChIP). Ten million cells were cross-linked for 10 min in 1% formaldehyde (Sigma) at room temperature, and quenched by adding glycine to a final concentration of 0.125 M for 5 min at room temperature. Cells were washed twice in chilled PBS, resuspended in buffer containing 50 mM HEPES-KOH (pH 7.5), 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100 and protease inhibitors (Thermo Scientific) and kept on ice for 10 min. Nuclei were pelleted at 800g for 5 min at 4 °C and resuspended in

buffer containing 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and protease inhibitors (Thermo Scientific) followed by a 10-min incubation on ice. Nuclei were collected and resuspended in sonication buffer containing 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% EGTA, 0.1% Na-deoxycholate, 0.5% *N*-lauryl sarcosine and protease inhibitors (Thermo Scientific). Sonication of DNA was performed in an ultra sonicator water bath (Bioruptor) using two ten cycle runs of 30 s 'on' and 30 s 'off' to achieve an average fragment length of 200–400 bp. After addition of 1% Triton X-100, samples were centrifuged at 16000g for 10 min at 4 °C. An aliquot of sonicated DNA was reverse-crosslinked and run on a 1% agarose gel to confirm fragment size during each ChIP procedure. Chromatin (25 µg) was immunoprecipitated by adding the antibody of interest followed by overnight incubation at 4 °C. The following antibodies were used for ChIP: anti-CTCF (Millipore 07-729, batch no. DAM1772428), anti-RNA polymerase II clone 4H8 (Millipore 05-623, batch no. DAM1731474), anti-Ser2P RNA polymerase II clone H5 (Covance MMS129R, batch no. E10017AF), anti-Ser5P RNA polymerase II clone H14 (Covance MMS134R, batch no. E10142DF), anti-H3K36Me3 (Abcam ab9050, batch no. 947467), anti-H3K27Me3 (Abcam ab6002, batch no. 934602), anti-H3K4Me3 clone MC315 (Millipore 04-745, batch no. NG1717145), Normal Rabbit IgG (Cell signaling Technology 2729, batch no. 4), and normal mouse IgG (Millipore 12-371, batch no. 1718089). After overnight incubation, 30 µl of Dynal Protein A/G beads (Invitrogen) or Protein L magnetic beads (Biovision) (for phosphorylated RNA Pol II antibodies) were added and incubated for 1 h at 4 °C. Beads were washed sequentially for 3 min each in low salt (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), high salt (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), LiCl buffer (10 mM Tris-HCl pH 8.0, 0.25 M LiCl, 1% NP40, 1% Na-deoxycholate) and TE buffer. Beads were eluted in 150 µl elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS, 50 mM NaHCO₃) and treated with 1 µl RNase A (1 mg ml⁻¹ Ambion) at 37 °C for 30 min. Cross-linking was reversed and proteins were degraded by addition of 1 µl proteinase K (20 mg ml⁻¹ Ambion) and incubation at 65 °C for 4 h. Eluted DNA was purified with QIAquick PCR purification (Qiagen), according to the manufacturer instructions.

Immunoprecipitated DNA and 5% input DNA were analysed by SYBR-Green real-time quantitative PCR. PCR measurements were performed in duplicate. The average cycle thresholds for the technical replicates were calculated to yield one value per primer set for each biological replicate and normalized to input using the formula $2^{(C_{\text{input}} - C_{\text{immunoprecipitation}})}$. These values were further normalized relative to the rabbit or mouse Ig control IP values for the primer set. Averages and standard deviations of the normalized biological replicate values were plotted in the figures and used in *t*-test calculations. Figure legends indicate the number of biological replicates (individual IPs) used in each experiment.

Methylated DNA immunoprecipitation (MedIP). MedIP was performed essentially according to the protocol described in ref. 44. Genomic DNA was purified from approximately 25 million cells using Zymo research Quick gDNA Midiprep kit (D3100), according to the manufacturer's instructions. For primary cells, CD3⁺ T cells were isolated from peripheral blood using CD3 microbeads (Miltenyi Biotech). CD3⁺ T cells were sorted into CD45RB high and CD45RB medium populations based on surface receptor staining of CD45RB and CD45RO. Purified genomic DNA was diluted into a total of 300 µl TE buffer and sonicated with a Bioruptor (10 cycles at low power, of 30 s 'on' and 30 s 'off') to an average size of 300–500 bp. An aliquot of sonicated DNA was run on 1% agarose gel to confirm fragment size during each MedIP procedure. Sonicated DNA (4 µg; 3 µg for primary cells) was denatured by incubation at 95 °C for 10 min and was immediately transferred to ice for 10 min. Immunoprecipitation buffer containing 10 mM sodium phosphate, 140 mM NaCl and 0.05% Triton X-100 was added to a final volume of 500 µl. For each IP reaction, 10 µg (8 µg for primary cells) of 5-methyl cytidine antibody clone b (Diagenode MAB-006-100, batch no. DA-0018) was added and incubated overnight at 4 °C with shaking. Five percent of DNA was kept as input.

After incubation, 30 µl of Dynal Protein G beads were added and further incubated for 1 h at 4 °C. Beads were washed thrice with 500 µl of IP buffer. Elution buffer (150 µl) containing 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS, 50 mM NaHCO₃ and 20 µg proteinase K was added and incubated at 55 °C for 3 h. Tubes were applied to a magnetic rack and eluted DNA and input DNA were purified with the QIAquick PCR purification kit (Qiagen) followed by SYBR-Green real-time quantitative PCR to identify methylated regions. PCR measurements were performed in duplicate. The average cycle thresholds for the technical replicates were calculated to yield one value per primer set for each biological replicate and normalized to input using the formula $2^{(C_{\text{input}} - C_{\text{immunoprecipitation}})}$. Averages and standard deviations of the normalized biological replicate values were plotted in the figures and used in *t*-test calculations. Figure legends indicate the number of biological replicates (individual IPs) used in each experiment.

In vitro transcription elongation assay. RNA pol II from yeast *Saccharomyces cerevisiae* containing a histidine-tagged Rpb3 subunit was purified as described previously⁴⁵. Histidine-tagged TFIIS expression plasmid⁴⁶ was a gift from C. Kane. Recombinant TFIIS was purified according to ref. 46, with an additional purification on a Mono-S column (GE Healthcare). Human CTCF recombinant protein was obtained from Abnova (catalogue no. H00010664-P01, batch no. 0991020-2).

Elongation complex incorporating a 9-nt RNA was assembled as described previously⁴⁷, purified with Amicon Ultra-0.5 ml centrifugal filter (Millipore), and diluted with transcription buffer (TB; 20 mM Tris-HCl pH 7.9, 5 mM MgCl₂, 10 mM 2-mercaptoethanol, 40 mM KCl, 0.1 mg ml⁻¹ BSA). The reaction was initiated by mixing 5 µl of TEC +/- X µM CTCF with 5 µl of 0.1–0.5 mM NTP (GE Healthcare) +/- 1 µM TFIIS in TB and was terminated with gel-loading buffer (5 M urea, 25 mM EDTA at final concentration). RNA products were resolved in 20% denaturing polyacrylamide gels and visualized with a Typhoon 8600 phosphorimager (GE Healthcare).

Oligonucleotides used for elongation complex. Sequences of RNA and DNA oligonucleotides are as follows. RNA, 5'-AUCGAGAGG-3'; DNA with CTCF binding site, non-template strand, 5'-GGTATAGGATACTTACAGCCATCGA GAGGGACAAGGCGAAAGCATCCACCAGGGGGCGCCAGCTAAT-3'; template strand, 5'-ATTAGCTGGCGCCCCCTGGTGGATGCTTTCGCTTGTCC CTCTCGATGGCTGTAAGTATCTTATACC-3'.

Electrophoretic mobility shift assay (EMSA). The CTCF-binding oligonucleotides used for EMSA correspond to either the template used for *in vitro* transcription (Supplementary Fig. 6a) or the CTCF binding sites in the wild-type and mutated I3-I7 minigenes (Supplementary Fig. 6d). The two strands of DNA were annealed, 5' end-labelled with [γ -³²P] ATP and purified with a G-50 Micro column (GE Healthcare). DNA probe (3 pM) equalling approximately 70,000 c.p.m. was mixed with glutathione S-transferase (GST)-tagged CTCF in binding buffer containing PBS and 5 mM MgCl₂, 0.1 mM ZnSO₄, 1 mM DTT, 0.1% NP40 and 10% glycerol. EMSA reaction mixtures (20 µl final volume) were incubated for 20 min at room temperature followed by electrophoresis on 5% native polyacrylamide gels and visualized as described above for *in vitro* transcription.

EMSA DNA probe sequences. *In vitro* transcription probe, 5'-CATCCACCAG GGGGCGCCAGCTAAT-3' and 5'-ATTAGCTGGCGCCCCCTGGTGGATG-3'; wild-type exon 5 probe, 5'-TCAGTTCAGCAGAGGGCGTCTGCG-3' and 5'-CGCAGACGCCCTCTGCTGGAAGTGA-3'; mutated exon 5 probe, 5'-TCAG TTAAGCTGAGTACGTCTGCG-3' and 5'-CGCAGACGTACTCAGCTTTAA CTGA-3'.

ChIP-Seq analyses. Illumina FastQ files were mapped to the human genome (hg19) using Bowtie⁴⁸ requiring a unique match (using the '-m 1' flag). The aligned reads in SAM format were converted to BED format before running the MACS peak caller⁴⁹. MACS peak calling was run using the rabbit Ig control sequencing data as background files for BJAB and BL41 CTCF ChIP-Seq data, respectively. The number of peaks identified per ChIP-Seq sample and sequenced reads are listed in Supplementary Table 2.

To investigate the effects of CTCF binding upon pre-mRNA splicing, we compared CTCF ChIP-Seq peaks with a set of alternative exons¹ requiring that the CTCF peak summit was located within 1,000 bp of the alternative exon boundaries (see Fig. 5a). Based on the presence of local CTCF peaks in BL41, BJAB and CD4 we classified each alternative exon into exons that were unbound by CTCF, and exons with either downstream or upstream CTCF binding. Classified unbound exons lacked CTCF peak summits in both the alternative exon body and within 1,000 bp on either side of the alternative exon. Exons with downstream CTCF binding had one or more CTCF peak summits within the region spanning from the alternative exon 5' splice site and 1,000 bp downstream in one or more of the CTCF data in BJAB, BL41 or CD4. Any alternative exons with a downstream CTCF peak but additional peak summits in the upstream region or within the alternative exon were not considered. The reciprocal procedure was used to classify exons with upstream CTCF binding. Alternative exons classified by local CTCF binding together with exon inclusion levels are provided in Supplementary Table 4.

RNA-Seq analyses. Illumina FastQ files were mapped to the human genome (hg19) and a collection of junctions using Tophat version 1.1.4 (ref. 50), using the paired-end mode and requiring a unique match. The resulting SAM file with uniquely mapped reads was converted to BAM format using samtools⁵¹. Mapping statistics for the RNA-Seq data are provided in Supplementary Table 3. We estimated exon inclusion levels of a collection of 42,557 alternative exons

(approximately the same as in ref. 1) using the MISO program³⁹ with the default parameters using the 'compute-genes-psi' function. The estimated exon inclusion levels from different RNA-Seq experiments were compared using MISO function 'compare-samples' to obtain exon inclusion level differences and Bayes factors. Statistically significant differences in exon inclusion levels between CTCF-bound and -unbound exons at different thresholds were evaluated using the Wilcoxon rank sum test. The overall difference in gene expression across RNA-Seq samples was evaluated using singular value decomposition. First we computed the expression level of each Refseq transcript as reads per kilobase and million mappable reads using the rpkmforgenes program⁵². The full gene expression matrix was normalized to unit length per transcript and subsequently used as input for singular value decomposition using the svdman program⁵³. The result in Supplementary Fig. 8c was obtained by projecting each sample onto the first two 'eigenarrays'⁵⁴.

RNA polymerase II ChIP-Seq analysis. We generated normalized RNA polymerase II fold enrichment signals over a set of regions by dividing the observed read sum at each position with the expected read sum computed as: (total_reads * read_length * number_of_regions) / genome_length. The normalized fold enrichment at each position was smoothed by window averaging using a window size of 100 nucleotides. We analysed all alternative exons with a downstream CTCF peak summit conserved in CD4⁺ T cells (Supplementary Table 4) after removing exons with a CTCF peak within 1,000 bp of an annotated transcript start site or poly A site in Ensembl (to remove effects from strong RNA pol II signals at transcript start and end locations). This procedure rendered 408 exons from which we computed both the CTCF peak summit position and exon middle coordinate. These two sets of 408 genomic coordinates each were used as the centre for the analysis in Fig. 5d. The same procedure was used to generate CTCF peak summits and middle exon positions for alternative exons with a conserved CTCF peak in the upstream region for Supplementary Fig. 8e (number of exons identified was 416).

Minigenes and transfection. CD45 minigenes were cloned into the pC1-neo mammalian expression vector (Promega). The wild-type CD45 minigene consists of 9.7 kb of CD45 genomic DNA sequence extending from 2.4 kb of intron 3 through 588 bp of intron 7. The I3-I7*CTCF minigene was made by mutating the CTCF binding site of exon 5 using site-directed mutagenesis with the primers indicated in Supplementary Table 1. The minigenes were transfected (Lipofectamine 2000, Invitrogen) along with pC1-neo vector control into HEK293 cells, CHO cells and NIH-3T3 cells. Cells were collected 48 h after transfection for RNA isolation (RNeasy, Qiagen) and chromatin immunoprecipitation. Transfection was performed in triplicate for HEK293 and CHO cells, with an individual RNA preparation (HEK293 and CHO) and duplicate ChIPs (CHO) derived from each of the three dishes. Transfection was performed in a single dish for NIH3T3 cells with three individual RNA preparations and triplicate ChIPs derived from the one dish.

44. Mohn, F., Weber, M., Schubeler, D. & Roloff, T. C. Methylated DNA immunoprecipitation (MeDIP). *Methods Mol. Biol.* **507**, 55–64 (2009).
45. Kireeva, M. L. *et al.* Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol. Cell* **30**, 557–566 (2008).
46. Awrey, D. E. *et al.* Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J. Biol. Chem.* **272**, 14747–14754 (1997).
47. Kireeva, M. L., Lubkowska, L., Komissarova, N. & Kashlev, M. Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from *Saccharomyces cerevisiae*. *Methods Enzymol.* **370**, 138–155 (2003).
48. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
49. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
50. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
51. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
52. Ramsköld, D., Wang, E. T., Burge, C. B. & Sandberg, R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* **5**, e1000598 (2009).
53. Wall, M. E., Dyck, P. A. & Brettin, T. S. SVDMAN—singular value decomposition analysis of microarray data. *Bioinformatics* **17**, 566–568 (2001).
54. Wall, M. E., Rechsteiner, A. & Rocha, L. M. in *A Practical Approach to Microarray Data Analysis* (eds Berrar, D. P., Dubitzky, W. & Granzow, M.) pp. 91–109 (Springer, 2003).

Low-Mach-number turbulence in interstellar gas revealed by radio polarization gradients

B. M. Gaensler¹, M. Haverkorn^{2,3,4}, B. Burkhart⁵, K. J. Newton-McGee^{1,6}, R. D. Ekers⁶, A. Lazarian⁵, N. M. McClure-Griffiths⁶, T. Robishaw¹, J. M. Dickey⁷ & A. J. Green¹

The interstellar medium of the Milky Way is multiphase¹, magnetized² and turbulent³. Turbulence in the interstellar medium produces a global cascade of random gas motions, spanning scales ranging from 100 parsecs to 1,000 kilometres (ref. 4). Fundamental parameters of interstellar turbulence such as the sonic Mach number (the speed of sound) have been difficult to determine, because observations have lacked the sensitivity and resolution to image the small-scale structure associated with turbulent motion^{5–7}. Observations of linear polarization and Faraday rotation in radio emission from the Milky Way have identified unusual polarized structures that often have no counterparts in the total radiation intensity or at other wavelengths^{8–12}, and whose physical significance has been unclear^{13–15}. Here we report that the gradient of the Stokes vector (Q , U), where Q and U are parameters describing the polarization state of radiation, provides an image of magnetized turbulence in diffuse, ionized gas, manifested as a complex filamentary web of discontinuities in gas density and magnetic field. Through comparison with simulations, we demonstrate that turbulence in the warm, ionized medium has a relatively low sonic

Mach number, $M_s \lesssim 2$. The development of statistical tools for the analysis of polarization gradients will allow accurate determinations of the Mach number, Reynolds number and magnetic field strength in interstellar turbulence over a wide range of conditions.

We consider radio-continuum images of an 18-deg² patch^{11,16} of the Galactic plane, observed with the Australia Telescope Compact Array (ATCA) at a frequency of 1.4 GHz. Data were simultaneously recorded in total intensity (Stokes parameter I) and in linear polarization (Stokes parameters Q and U). The Stokes I image (Fig. 1) shows a typical distribution of radio emission, consisting of supernova-remnant shells, ionized regions around massive stars (H II regions) and unresolved distant radio sources. However, the corresponding images of Q , U and the linearly polarized intensity $P \equiv (Q^2 + U^2)^{1/2}$ in Fig. 1 are filled with complex structure that bears little resemblance to the Stokes I image, as has also been seen in many other polarimetric observations at radio frequencies^{8,9,12}. The intensity variations seen in Q , U and P are the result of small-scale angular structure in the Faraday rotation induced by ionized gas⁸, and are thus an indirect representation of

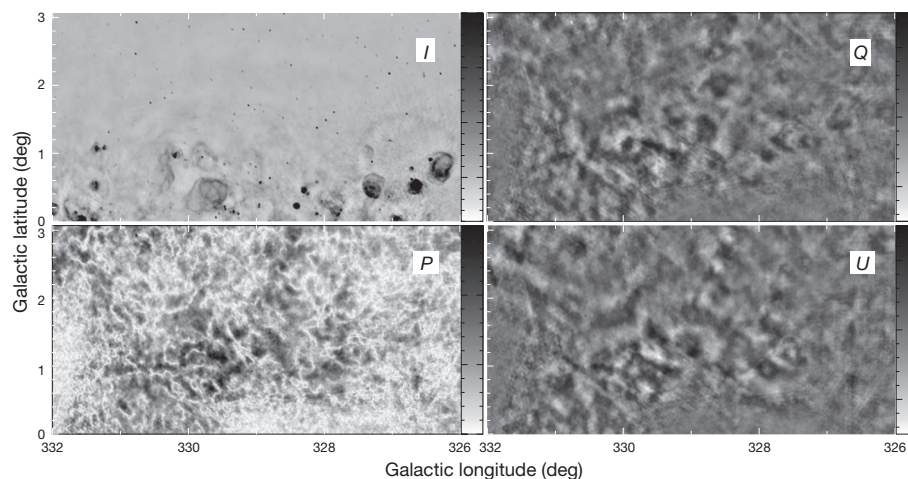


Figure 1 | Total intensity (I) and linearly polarized intensity (Q , U , P) for an 18-deg² region of the Southern Galactic Plane Survey²⁹. All four images were generated¹¹ from a set of observations¹⁶ taken at the ATCA over the period 1997 April to 1998 April using a 96-MHz bandwidth centred on an observing frequency of 1,384 MHz. The field is a mosaic of 190 pointings each with a total integration time of 20 min, resulting in an approximately uniform sensitivity, over most of the field, of 0.8 mJy per beam (Stokes I) or 0.55 mJy per beam (Stokes Q and U) at an angular resolution of 75 arcsec ($1 \text{ Jy} \equiv 10^{-26} \text{ W m}^{-2} \text{ Hz}^{-1}$). The scale for each image is shown on the right of each panel. The Stokes I image is displayed over a range of -40 to $+150$ mJy per beam (each interval corresponds to 10 mJy per beam). Because the ATCA is an interferometer, it is

not sensitive to structure on angular scales larger than 35 arcmin. Faint wisps can be seen, corresponding to the sharp edges of large-scale structures. However, the bulk of the smooth radio emission from Galactic cosmic rays is not detected. Imaging artefacts in the form of grating rings and radial streaks can be seen around a few very bright sources, but these regions were not used in our statistical analysis. The Stokes Q and U images are displayed over a range of -15 to $+15$ mJy per beam (interval, 2 mJy per beam), and the P image covers a range of 0 to 15 mJy per beam (interval, 1 mJy per beam). Almost none of the structure seen in Q , U and P has any correspondence with any emission seen in Stokes I ; the mottled structure results from spatial fluctuations in Faraday rotation in the ISM.

¹Sydney Institute for Astronomy, School of Physics, The University of Sydney, New South Wales 2006, Australia. ²ASTRON, Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands. ³Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands. ⁴Department of Astrophysics/IMAPP, Radboud University Nijmegen, PO Box 9010, 6500 GL Nijmegen, The Netherlands. ⁵Astronomy Department, University of Wisconsin, Madison, 475 North Charter Street, Madison, Wisconsin 53711, USA. ⁶Australia Telescope National Facility, CSIRO Astronomy and Space Science, PO Box 76, Epping, New South Wales 1710, Australia. ⁷School of Mathematics and Physics, University of Tasmania, Private Bag 37, Hobart, Tasmania 7001, Australia.

turbulent fluctuations in the free-electron density and magnetic field throughout the interstellar medium¹⁵ (ISM).

A limitation of previous studies is that they usually interpreted the data in terms of the amplitude, P , and/or the angle, $\theta \equiv (1/2)\tan^{-1}(U/Q)$, of the complex Stokes vector $\mathbf{P} \equiv (Q, U)$. However, neither polarization amplitude nor polarization angle is preserved under arbitrary translations and rotations in the Q – U plane. These can result from one or more of a smooth distribution of intervening polarized emission, a uniform screen of foreground Faraday rotation, and the effects of missing large-scale structure in an interferometric data set. In the most general case, we are thus forced to conclude that the observed values of P and θ do not have any physical significance, and that only measurements of quantities that are both translationally and rotationally invariant in the Q – U plane can provide insight into the physical conditions that produce the observed polarization distribution.

The simplest such quantity is the spatial gradient of \mathbf{P} , that is, the rate at which the polarization vector traces out a trajectory in the Q – U plane as a function of position on the sky. The magnitude of the gradient is unaffected by rotation and translation, and so has the potential to reveal properties of the polarization distribution that might otherwise be hidden by excess foreground emission or Faraday rotation, or in data sets from which large-scale structure is missing (as is the case for the data shown in Fig. 1). The magnitude of the polarization gradient is

$$|\nabla\mathbf{P}| = \sqrt{\left(\frac{\partial Q}{\partial x}\right)^2 + \left(\frac{\partial U}{\partial x}\right)^2 + \left(\frac{\partial Q}{\partial y}\right)^2 + \left(\frac{\partial U}{\partial y}\right)^2} \quad (1)$$

The expression in equation (1) can be calculated simply, and the corresponding image of $|\nabla\mathbf{P}|$ (Fig. 2) reveals a complex network of tangled filaments. In particular, all regions in which $|\nabla\mathbf{P}|$ is high consist of elongated, narrow structures rather than extended patches. In the inset of Fig. 2, we plot the direction of $\nabla\mathbf{P}$ for a small subregion of the image, demonstrating that $\nabla\mathbf{P}$ changes most rapidly along directions oriented perpendicular to the filaments. We can explore the

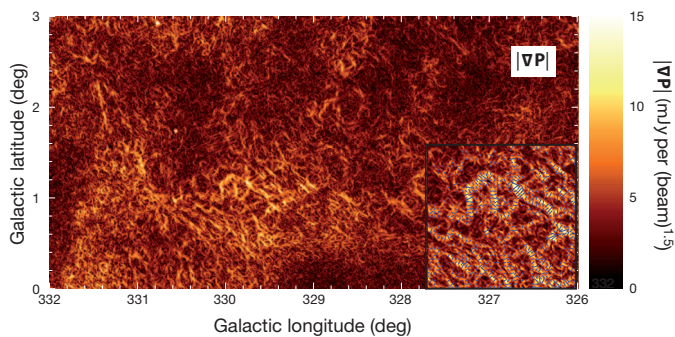


Figure 2 $|\nabla\mathbf{P}|$ for an 18-deg² region of the Southern Galactic Plane Survey. $|\nabla\mathbf{P}|$ has been derived by applying equation (1) to the Q and U images from Fig. 1; note that $|\nabla\mathbf{P}|$ cannot be constructed from the scalar quantity $P \equiv (Q^2 + U^2)^{1/2}$, but is derived from the vector field $\mathbf{P} \equiv (Q, U)$. $|\nabla\mathbf{P}|$ is a gradient in one dimension, for which the appropriate units are (beam)^{−0.5}. Because P measures linearly polarized intensity in units of millijanskys per beam, $|\nabla\mathbf{P}|$ has units of millijanskys per (beam)^{1.5}. The scale showing $|\nabla\mathbf{P}|$ is shown on the right of the image, and ranges from 0 to 15 mJy per (beam)^{1.5}. The inset shows an expanded version of the structure with highest $|\nabla\mathbf{P}|$, covering a box of side 0.9 deg centred on Galactic longitude 329.8 deg and Galactic latitude +1.0 deg. Plotted in the inset is the direction of $\nabla\mathbf{P}$ at each position, defined as

$$\arg(\nabla\mathbf{P}) \equiv \tan^{-1} \left[\text{sign} \left(\frac{\partial Q}{\partial x} \frac{\partial Q}{\partial y} + \frac{\partial U}{\partial x} \frac{\partial U}{\partial y} \right) \sqrt{\left(\frac{\partial Q}{\partial y}\right)^2 + \left(\frac{\partial U}{\partial y}\right)^2} / \sqrt{\left(\frac{\partial Q}{\partial x}\right)^2 + \left(\frac{\partial U}{\partial x}\right)^2} \right]$$

For clarity, vectors are shown only at points where the amplitude of the gradient is greater than 5 mJy per (beam)^{1.5}.

frequency dependence of these filaments, because the 1.4-GHz ATCA data shown in Fig. 1 consist of nine independent spectral channels of width 8 MHz, spread over a total bandwidth of 96 MHz. We have constructed images of $|\nabla\mathbf{P}|$ for each individual spectral channel, and these show the same set of specific features as in the overall image, albeit at reduced signal-to-noise ratios. The lack of frequency dependence indicates that the high-gradient structures seen in this data set correspond to physical features in the ISM rather than to contour lines introduced by the particular combination of observing frequency and angular resolution used^{15,17}.

We first consider the possibility that these filaments of high gradient are intrinsic to the source of emission. Abrupt spatial transitions in the strength or geometry of the magnetic field in a synchrotron-emitting region would generate a large gradient in (Q, U) . However, processes of that sort would also produce structure in the overall synchrotron emissivity, such that we would observe features in the image of Stokes I that match those seen in $|\nabla\mathbf{P}|$. No such correspondence is observed, demonstrating that the regions of high polarization gradient are not intrinsic to the source of polarized emission but must be induced by Faraday rotation in magneto-ionized gas.

Because the amount of Faraday rotation is proportional to the line integral of $n_e B_{\parallel}$ from the source to the observer (where n_e is the density of free electrons and B_{\parallel} is the uniform component of the line-of-sight magnetic field), the filamentary structure seen in $|\nabla\mathbf{P}|$ must correspond to boundaries across which n_e and/or B_{\parallel} show a sudden increase or decrease over a small spatial interval. Such discontinuities could be shock fronts or ionization fronts from discrete sources, as have been observed in polarization around the rims of supernova remnants, H II regions and planetary nebulae^{11,18}. We have examined this possibility by carefully comparing our image of $|\nabla\mathbf{P}|$ with images and gradient images of Stokes I (tracing shock waves seen in synchrotron emission; ref. 11), 21-cm H I emission¹⁶ (tracing atomic hydrogen) and 656.3-nm H α emission^{19,20} (tracing ionized hydrogen) over the same field, but do not find any correspondences.

We conclude that the features seen in $|\nabla\mathbf{P}|$ are a generic component of diffuse, ionized gas in this direction in the sky. To test this hypothesis, we performed a series of three-dimensional isothermal simulations of magnetohydrodynamic turbulence in the ISM, each with different parameters for the sonic Mach number, defined as $M_s \equiv \langle |\mathbf{v}| / c_s \rangle$, where \mathbf{v} is the local velocity, c_s is the sound speed and the averaging (indicated by angle brackets) is done over the whole simulation. For each simulation, we propagated a uniform source of polarized emission through the distribution of turbulent, magnetized gas. The resultant Faraday rotation produces a complicated distribution on the sky of Stokes Q and U , from which we generated a map of the polarization gradient using equation (1). Images of $|\nabla\mathbf{P}|$ for representative simulations of the subsonic, transonic and supersonic regimes are shown in Fig. 3. Narrow, elongated filaments of high polarization gradient are apparent in each simulation in Fig. 3, although they differ in their morphology and degree of organization. In particular, the supersonic case (Fig. 3c) shows localized groupings of very high-gradient filaments, corresponding to ensembles of intersecting shocks^{5,21,22}. By contrast, the subsonic (Fig. 3a) and transonic (Fig. 3b) cases show more-diffuse networks of filaments, representing the cusps and discontinuities characteristic of any turbulent velocity field^{6,21,23}.

Visual comparison of the simulated distributions of $|\nabla\mathbf{P}|$ with real data (Fig. 2) suggests that the subsonic and transonic cases shown in Fig. 3a, b more closely resemble the observations than does the supersonic case. We can quantify this statement by calculating the third-order moment (skew, γ) and the fourth-order moment (kurtosis, β) of the probability distribution function of $|\nabla\mathbf{P}|$ for both observations and simulations: these quantities parameterize the degree of Gaussian asymmetry in the probability distribution function, and hence provide information on the amount of compression due to shocks in the data^{6,24}.

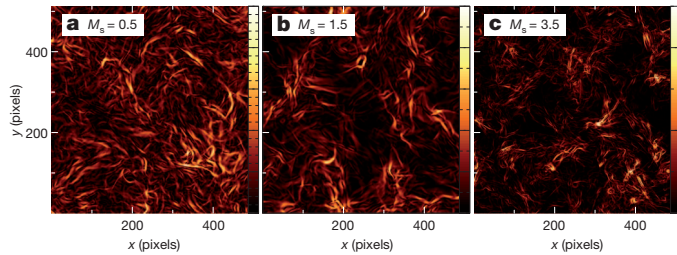


Figure 3 $|\nabla P|$ derived from propagation of linear radio polarization through three different isothermal simulations of magnetized turbulence. Each simulation is a $512 \times 512 \times 512$ -element periodic box with a linear dimension of 0.15 pc for each pixel, evolved in time using an essentially non-oscillatory scheme^{24,30}. Three such simulations are shown, each labelled with its corresponding value for M_s : subsonic ($M_s < 1$; **a**), transonic ($M_s \approx 1$; **b**) and supersonic ($M_s > 1$; **c**). At the start of each simulation, the electron density had a uniform value $n_e = 0.1 \text{ cm}^{-3}$ and the magnetic field was oriented in the plane of the sky with a uniform amplitude of $B = 0.3 \text{ } \mu\text{G}$ (subsonic), $1 \text{ } \mu\text{G}$ (transonic) or $2 \text{ } \mu\text{G}$ (supersonic), which corresponds to a constant Alfvénic Mach number of $M_A = 2$ in each case. Turbulence was driven solenoidally in Fourier space at large scales (small wavenumber) until the turbulent cascade had fully developed and a steady state between input energy and dissipation had been reached. In each case, we illuminated the simulation volume with a background radio source of uniform polarization at an emission frequency of 1.4 GHz, with $Q = 100 \text{ mJy}$ per pixel and $U = 0 \text{ mJy}$ per pixel at all positions. At each pixel, the line integral of $n_e B_{\parallel}$ was computed, and the corresponding Faraday rotation was applied to the polarized signal, to calculate values of Q and U . No effects due to finite angular resolution, depolarization or incomplete interferometric visibility coverage were included, so the observed polarized signal is $P = 100 \text{ mJy}$ per pixel at all positions. We then calculated the gradient, $|\nabla P|$, using equation (1). The scales showing $|\nabla P|$ are shown on the right of the images, and range from 0 to 25 mJy per (pixel)^{1.5} for **a** (interval, 1 mJy per (pixel)^{1.5}), 0 to 100 mJy per (pixel)^{1.5} for **b** (interval, 10 mJy per (pixel)^{1.5}) and 0 to 500 mJy per (pixel)^{1.5} for **c** (interval, 100 mJy per (pixel)^{1.5}).

In the simulations, we found that both the skew and the kurtosis of $|\nabla P|$ increased monotonically with sonic Mach number. We used a genetic algorithm²⁵ to determine that the threshold for strongly supersonic turbulence was $\gamma > 1.1$ and $\beta > 1.5$. We then computed the third- and fourth-order moments for the observed distribution of $|\nabla P|$ shown in Fig. 2, and found that $\gamma = 0.3$ and $\beta = 0.9$.

This analysis of the moments of the polarization gradient therefore confirms quantitatively what we concluded above from visual inspection: the turbulent, ionized ISM in this direction in the sky is subsonic or transonic. The findings we obtained by imaging the polarization gradients produced by interstellar turbulence are supported by recent statistical studies of H α emission measures and of 21-cm H I column densities over large volumes, which have similarly found that $M_s \lesssim 2$ for warm gas throughout the ISM^{26,27}.

In the simulations shown in Fig. 3, the sharp gradients in (Q, U) occur as a result of localized high values of the gas density and magnetic field, resulting from vorticity or shock compression. However, the filamentary features seen in $|\nabla P|$ may not be easily observable in other types of data: for example, if we adopt typical parameters for warm, ionized gas^{1,28} of $n_e \approx 0.3 \text{ cm}^{-3}$ and $B_{\parallel} \approx 2 \text{ } \mu\text{G}$, even the compression associated with a strong adiabatic shock produces across-filament changes in emission measure and Faraday rotation measure of only $\sim 0.5 \text{ pc cm}^{-6}$ and $\lesssim 5 \text{ rad m}^{-2}$, respectively, assuming a spatial scale^{7,14} for these structures of $\sim 0.5 \text{ pc}$. This is below observable levels in H α and other tracers of emission measure. The rotation measure gradient¹⁴ across these interfaces is potentially observable in spectropolarimetric radio data, but the addition of single-dish observations is required to recover the total power of the polarized signal. By contrast, even a small gradient in rotation measure can produce an arbitrarily large value of $|\nabla P|$ (irrespective of whether single-dish measurements are present in the data), provided that there is a strong source of background polarized emission through which the discontinuities in Faraday rotation are viewed. Further investigation of the polarization

gradient and its statistical properties will provide robust estimates of poorly constrained parameters of turbulent flows such as the sonic and Alfvénic Mach numbers, the characteristic magnetic field strength, the Reynolds number and the physical scale of energy injection.

Received 3 March; accepted 11 August 2011.

Published online 5 October 2011.

1. Ferrière, K. M. The interstellar environment of our galaxy. *Rev. Mod. Phys.* **73**, 1031–1066 (2001).
2. de Avillez, M. A. & Breitschwerdt, D. Global dynamical evolution of the ISM in star forming galaxies. I. High resolution 3D simulations: effect of the magnetic field. *Astron. Astrophys.* **436**, 585–600 (2005).
3. Chepurnov, A. & Lazarian, A. Extending the big power law in the sky with turbulence spectra from Wisconsin H α Mapper data. *Astrophys. J.* **710**, 853–858 (2010).
4. Armstrong, J. W., Rickett, B. J. & Spangler, S. R. Electron density power spectrum in the local interstellar medium. *Astrophys. J.* **443**, 209–221 (1995).
5. Kritsuk, A. G., Norman, M. L., Padoan, P. & Wagner, R. The statistics of supersonic isothermal turbulence. *Astrophys. J.* **665**, 416–431 (2007).
6. Kowal, G., Lazarian, A. & Beresnyak, A. Density fluctuations in MHD turbulence: Spectra, intermittency, and topology. *Astrophys. J.* **658**, 423–445 (2007).
7. Kissmann, R., Kleimann, J., Fichtner, H. & Grauer, R. Local turbulence simulations for the multiphase ISM. *Mon. Not. R. Astron. Soc.* **391**, 1577–1588 (2008).
8. Wieringa, M. H., de Bruyn, A. G., Jansen, D., Brouw, W. N. & Katgert, P. Small scale polarization structure in the diffuse galactic emission at 325 MHz. *Astron. Astrophys.* **268**, 215–229 (1993).
9. Gray, A. D., Landecker, T. L., Dewdney, P. E. & Taylor, A. R. A large-scale, interstellar Faraday-rotation feature of unknown origin. *Nature* **393**, 660–662 (1998).
10. Haverkorn, M., Katgert, P. & de Bruyn, A. G. Structure in the local Galactic ISM on scales down to 1 pc, from multi-band radio polarization observations. *Astron. Astrophys.* **356**, L13–L16 (2000).
11. Gaensler, B. M. et al. Radio polarization from the inner galaxy at arcminute resolution. *Astrophys. J.* **549**, 959–978 (2001).
12. Uyaniker, B. & Landecker, T. L. A highly ordered Faraday rotation structure in the interstellar medium. *Astrophys. J.* **575**, 225–233 (2002).
13. Shukurov, A. & Berkhuisen, E. M. Faraday ghosts: depolarization canals in the galactic radio emission. *Mon. Not. R. Astron. Soc.* **342**, 496–500 (2003).
14. Haverkorn, M. & Heitsch, F. Canals beyond Mars: beam depolarization in radio continuum maps of the warm ISM. *Astron. Astrophys.* **421**, 1011–1019 (2004).
15. Fletcher, A. & Shukurov, A. Depolarization canals and interstellar turbulence. *EAS Publ. Ser.* **23**, 109–128 (2007).
16. McClure-Griffiths, N. M. et al. The Southern Galactic Plane Survey: the test region. *Astrophys. J.* **551**, 394–412 (2001).
17. Newton-McGee, K. J. *Radio Polarimetry as a Probe of Interstellar Magnetism* 83–87. Ph.D. thesis, Univ. Sydney (2009).
18. Ransom, R. R., Uyaniker, B., Kothes, R. & Landecker, T. L. Probing the magnetized interstellar medium surrounding the planetary nebula Sh 2–216. *Astrophys. J.* **684**, 1009–1017 (2008).
19. Gaustad, J. E., McCullough, P. R., Rosing, W. & Van Buren, D. A robotic wide-angle H α survey of the southern sky. *Publ. Astron. Soc. Pacif.* **113**, 1326–1348 (2001).
20. Parker, Q. A. et al. The AAO/UKST SuperCOSMOS H α survey. *Mon. Not. R. Astron. Soc.* **362**, 689–710 (2005).
21. Beresnyak, A., Lazarian, A. & Cho, J. Density scaling and anisotropy in supersonic magnetohydrodynamic turbulence. *Astrophys. J.* **624**, L93–L96 (2005).
22. Lemaster, M. N. & Stone, J. M. Dissipation and heating in supersonic hydrodynamic and MHD turbulence. *Astrophys. J.* **691**, 1092–1108 (2009).
23. Eyink, G. L. Besov spaces and the multifractal hypothesis. *J. Stat. Phys.* **78**, 353–375 (1995).
24. Burkhardt, B., Falceta-Gonçalves, D., Kowal, G. & Lazarian, A. Density studies of MHD interstellar turbulence: statistical moments, correlations and bispectrum. *Astrophys. J.* **693**, 250–266 (2009).
25. Whitley, D. A genetic algorithm tutorial. *Stat. Comput.* **4**, 65–85 (1994).
26. Hill, A. S. et al. The turbulent warm ionized medium: emission measure distribution and MHD simulations. *Astrophys. J.* **686**, 363–378 (2008).
27. Burkhardt, B., Stanimirović, S., Lazarian, A. & Kowal, G. Characterizing magnetohydrodynamic turbulence in the Small Magellanic Cloud. *Astrophys. J.* **708**, 1204–1220 (2010).
28. Gaensler, B. M., Madsen, G. J., Chatterjee, S. & Mao, S. A. The vertical structure of warm ionised gas in the Milky Way. *Publ. Astron. Soc. Aust.* **25**, 184–200 (2008).
29. Haverkorn, M., Gaensler, B. M., McClure-Griffiths, N. M., Dickey, J. M. & Green, A. J. The Southern Galactic Plane Survey: polarized radio continuum observations and analysis. *Astrophys. J.* **167** (suppl.), 230–238 (2006).
30. Cho, J. & Lazarian, A. Compressible magnetohydrodynamic turbulence: mode coupling, scaling relations, anisotropy, viscosity-damped regime and astrophysical implications. *Mon. Not. R. Astron. Soc.* **345**, 325–339 (2003).

Acknowledgements We thank S. Brown, A. Hill, R. Kissmann, A. MacFadyen, M.-M. Mac Low, E. Petroff, P. Slane and X. Sun for discussions. The Australia Telescope Compact Array is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO. B.M.G. and T.R. acknowledge the support of the Australian Research Council through grants FF0561298, FL100100114 and FS100100033. B.B. acknowledges support from the National Science Foundation Graduate Research Fellowship and the NASA Wisconsin Space Grant Institution. A.L. acknowledges the support of the National Science Foundation through grant AST0808118 and of the

Center for Magnetic Self-Organization in Astrophysical and Laboratory Plasmas. We thank the staff of the Australia Telescope National Facility, especially M. Calabretta, R. Haynes, D. McConnell, J. Reynolds, R. Sault, R. Wark and M. Wieringa, for their support of the Southern Galactic Plane Survey.

Author Contributions J.M.D., N.M.Mc.-G., B.M.G. and A.J.G. carried out the original observations. B.M.G., N.M.Mc.-G. and T.R. produced the polarization images from the raw data. B.M.G., M.H., K.J.N.-Mc., R.D.E and N.M.Mc.-G. worked together to develop the gradient technique, and B.M.G. then applied the gradient technique to the images. B.B.

and A.L. performed the simulations and the statistical analysis. B.M.G. led the writing of the paper and the interpretation of results. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.M.G. (bryan.gaensler@sydney.edu.au).

Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT

Nien-Jen Hu^{1,2,3}, So Iwata^{1,2,3,4,5}, Alexander D. Cameron^{1,2,3,4} & David Drew¹

High cholesterol levels greatly increase the risk of cardiovascular disease. About 50 per cent of cholesterol is eliminated from the body by its conversion into bile acids. However, bile acids released from the bile duct are constantly recycled, being reabsorbed in the intestine by the apical sodium-dependent bile acid transporter (ASBT, also known as SLC10A2). It has been shown in animal models that plasma cholesterol levels are considerably lowered by specific inhibitors of ASBT^{1,2}, and ASBT is thus a target for hypercholesterolaemia drugs. Here we report the crystal structure of a bacterial homologue of ASBT from *Neisseria meningitidis* (ASBT_{NM}) at 2.2 Å. ASBT_{NM} contains two inverted structural repeats of five transmembrane helices. A core domain of six helices harbours two sodium ions, and the remaining four helices pack in a row to form a flat, 'panel'-like domain. Overall, the architecture of the protein is remarkably similar to the sodium/proton antiporter NhaA³, despite having no detectable sequence homology. The ASBT_{NM} structure was captured with the substrate taurocholate present, bound between the core and panel domains in a large, inward-facing, hydrophobic cavity. Residues near this cavity have been shown to affect the binding of specific inhibitors of human ASBT⁴. The position of the taurocholate molecule, together with the molecular architecture, suggests the rudiments of a possible transport mechanism.

ASBT is an SLC10 (sodium bile acid co-transporter family) member that moves bile acids across the apical membrane of the ileum into the portal blood vein^{5,6}. ASBT uses the sodium ion gradient to drive the 'uphill' transport of bile acids across membranes, with a reported stoichiometry of two sodium ions per substrate⁷. Mutations in the human ASBT gene cause a condition of primary bile acid malabsorption⁸. ASBT is a pharmaceutical target for drugs aimed at lowering cholesterol, and several ASBT inhibitors have been developed that are effective in animal models^{1,2}. Because some drugs are poorly absorbed in the intestine or need to be targeted to the liver, ASBT and its close liver paralogue, NTCP (SLC10A1), have also received attention as prodrug carriers, capable of transporting various compounds coupled to bile acid, for example HMG-CoA reductase (HMGCR) inhibitors, the antiviral drug acyclovir, nucleotides and cytostatic drugs⁹.

ASBT_{NM} from *N. meningitidis*, with 26% identity and 54% similarity to human ASBT, was identified by fluorescence-based screening methods^{10,11} as a suitable candidate for structural studies (Supplementary Figs 1 and 2). Residues known to be functionally important in mammalian ASBT and other SLC10 members¹² are well conserved in ASBT_{NM} (Supplementary Fig. 1). Bile acid transport by ASBT_{NM} was confirmed in whole cells by the sodium-dependent uptake of [³H]-taurocholate (Fig. 1a). The observed Michaelis constant, K_m , for [³H]-taurocholate is ~50 μM (Fig. 1b), a value similar to that measured for rat and human ASBT^{7,13,14}. The ASBT inhibitors cyclosporin A¹⁵ and bromosulphophthalein¹⁵ and the drug fluvastatin¹⁶ are also competitors for ASBT_{NM}-mediated [³H]-taurocholate transport (Fig. 1c). Thus, ASBT_{NM} is a valid model of mammalian bile acid transporters. We

solved the structure of ASBT_{NM} by single-wavelength anomalous scattering and refined it at a resolution of 2.2 Å (Supplementary Tables 1 and 2; Methods).

ASBT_{NM} has cytoplasmic amino and carboxy termini, comprises ten transmembrane helices (TMs) that are linked by short loops, and has overall dimensions of approximately 45 Å × 30 Å × 30 Å (Fig. 2 and Supplementary Fig. 3). TM1 to TM5 and TM6 to TM10 are topologically similar but oppositely orientated in the plane of the

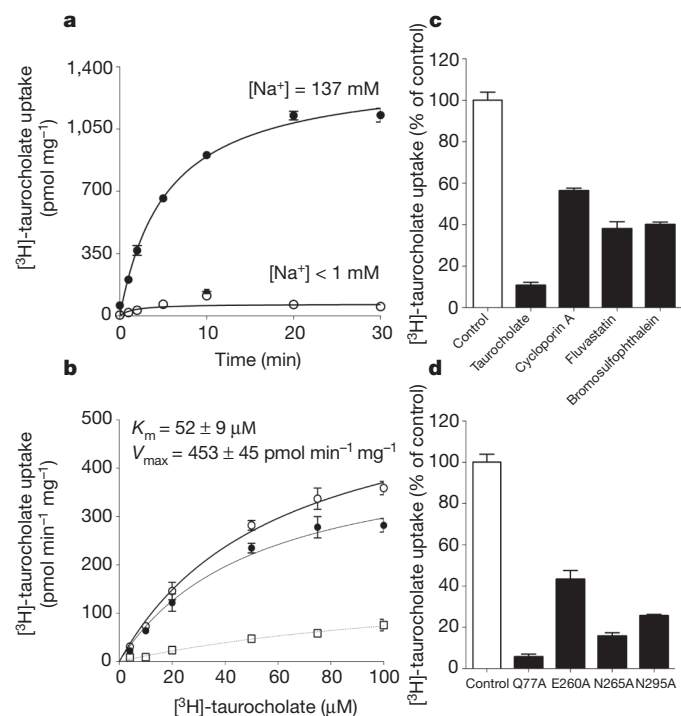


Figure 1 | Sodium-dependent transport of bile acid by ASBT_{NM}. **a**, Time-dependent uptake of [³H]-taurocholate after expression of ASBT_{NM} in *Escherichia coli*, as monitored in buffer containing 137 mM sodium (filled circles) or <1 mM sodium (open circles) **b**, Michaelis-Menten transport kinetics of ASBT_{NM}-mediated [³H]-taurocholate uptake. The specific uptake (filled circles) was calculated by subtracting the internalization measured from control cells lacking the transporter (open squares) from the total uptake (open circles), as detailed in Methods. **c**, ASBT_{NM}-mediated [³H]-taurocholate uptake after 5 min in the presence of 150 μM taurocholate, cyclosporin A, fluvastatin or bromosulphophthalein (black bars), measured as a percentage of the uptake without their addition (white bar). **d**, ASBT_{NM}-mediated [³H]-taurocholate uptake after 5 min for a wild-type control (white bar) and the indicated single alanine point mutants (black bars). The uptake for the mutants is displayed as a percentage of the wild-type activity. The expression and detergent-solubilized folded state of each mutant was similar to wild-type protein (Supplementary Fig. 2a). Errors bars, s.e.m.; $n = 3$.

¹Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, UK. ²Membrane Protein Laboratory, Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Chilton OX11 0DE, UK. ³Research Complex at Harwell Rutherford, Appleton Laboratory, Harwell, Oxford, Didcot OX11 0FA, UK. ⁴Japan Science and Technology Agency, ERATO, Human Crystallography Project, Yoshida Konoe, Sakyo-ku, Kyoto 606-851, Japan. ⁵Department of Cell Biology, Graduate School of Medicine, Kyoto University, Yoshida Konoe, Sakyo-ku, Kyoto 606-8501, Japan.

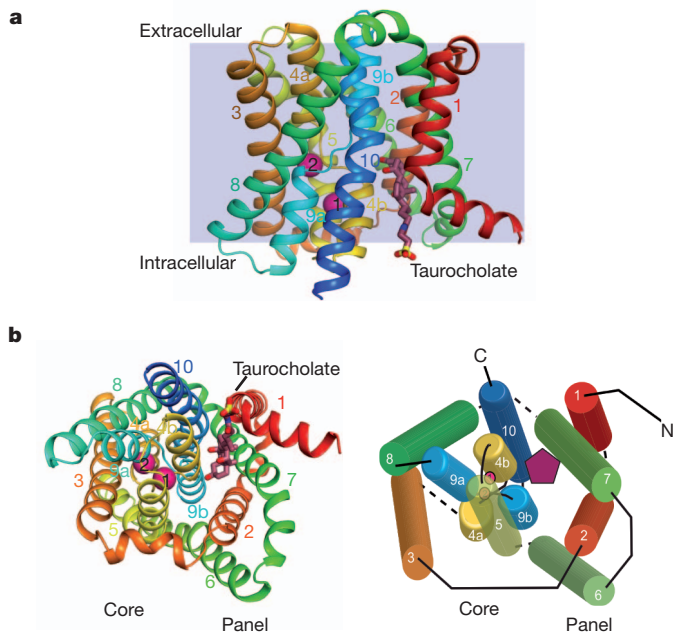


Figure 2 | ASBT_{NM} structure. **a**, Ribbon representation of ASBT_{NM} as viewed in the plane of the membrane. TM1 to TM10 have been coloured from red at the N terminus to blue at the C terminus, and the position of the membrane is depicted in grey. The pink spheres indicate sodium sites Na1 and Na2, and the stick model represents the substrate taurocholate. **b**, ASBT_{NM} structure as viewed from the intracellular side as a ribbon representation (left) and as a simplified cartoon (right). On the right, the burgundy pentagon represents taurocholate.

membrane. The root mean squared deviation (r.m.s.d.) after superposition of the two topology-inverted repeats is 3.7 Å (Supplementary Fig. 4a, b; Methods). Each repeating unit is made of an N-terminal V-motif (TM1 and TM2, TM6 and TM7) and a core motif of three helices (TM3 to TM5, TM8 to TM10) (Fig. 2 and Supplementary Figs 3 and 4). If the V-motif and the core motif are superposed separately, the r.m.s.d. in each case is lower: 2.6 Å and 2.8 Å, respectively (Supplementary Fig. 4c). The core motifs from each repeat form the core domain, whereas the two V-motifs create a panel-like domain (Fig. 2b). TM4 and TM9 in the core domain are broken in the middle (discontinuous) and form helical hairpins with TM5 and TM10, respectively, which are both kinked. At the point where TM4 and TM9 are broken by well-conserved peptide motifs, they cross over (Fig. 2 and Supplementary Figs 5 and 6). On the intracellular side, a wide crevice separates the core domain from the panel domain (Fig. 3a). The cavity extends over halfway through the protein. The extracellular side of the cavity is tightly closed by TM1, TM2, TM4b, TM7, TM9b and TM10. Previously, two topology models of ASBT were proposed with seven and nine transmembrane helices, respectively^{17,18}. Because TM1 is not conserved in ASBT, the structure is broadly consistent with the model with nine transmembrane helices (Supplementary Fig. 5). TM4 and TM9 were annotated as extracellular loops in the topology model with seven transmembrane helices, but were correctly identified in the model with nine transmembrane helices.

Discontinuous transmembrane helices are a common motif in secondary active transporters^{3,19,20}. However, the sodium/proton antiporter NhaA is the only other known example in which these helices cross as observed in ASBT_{NM} (Supplementary Fig. 6). Indeed, ASBT_{NM} has a similar structure to NhaA, and they superpose with an r.m.s.d. of 2.9 Å over 202 Cα atoms (Supplementary Fig. 7a; Methods). The similarity is more striking when the core and panel domains are superposed separately (Supplementary Fig. 7b). This unexpected finding further emphasizes the remarkable plasticity of transporters, which allows them to use a common scaffold to translocate different substrates²⁰.

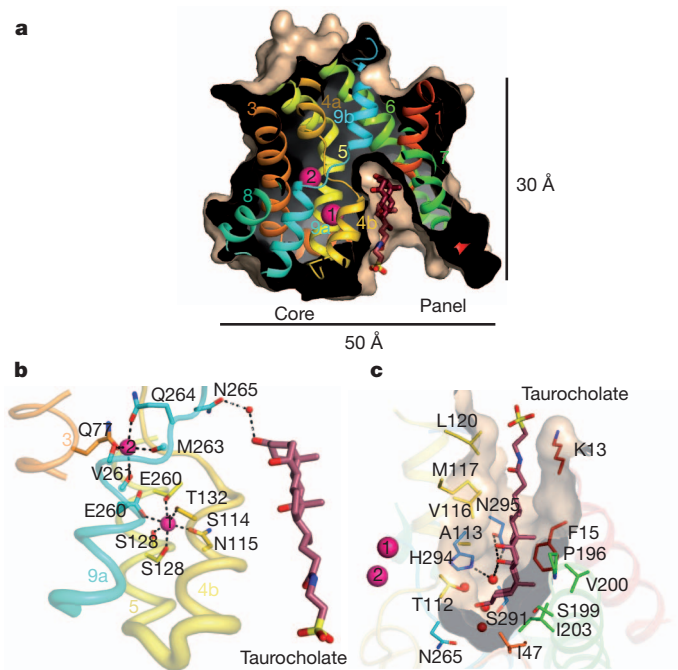


Figure 3 | ASBT_{NM} structure is inward facing and contains bound sodium and bile acid. **a**, Surface representation showing the location of the intracellular cavity in which taurocholate binds, as a section through the protein. **b**, The sodium-binding sites in ASBT_{NM}. Na1 is octahedrally coordinated by Ser 114 and Asn 115 on TM4b, Thr 132 and Ser 128 on TM5, and Glu 260 on TM9a. The square pyramidal arrangement of the Na2 ligands is made up of Glu 260, Val 261, Met 263 and Gln 264 on TM9 and Gln 77 on TM3. **c**, The intracellular cavity in ASBT_{NM}. Residues lining the cavity and near to the taurocholate are shown. The figures have been coloured as in Fig. 2. A 150-fold difference in inhibition of the mouse and human forms of ASBT by benzothiazepines⁴ has been ascribed to sequence differences corresponding to Ser 291 at the bottom of the cavity. Supplementary Fig. 10 shows stereo versions of **b** and **c**.

In ASBT and NTCP, two sodium ions are translocated per bile acid molecule^{7,21}. In the highly conserved core domain of ASBT_{NM} (Supplementary Fig. 8), we have identified two sodium-binding sites (Na1 and Na2) on the basis of the coordination and bond distances (2.0–2.5 Å) (Fig. 3b and Supplementary Figs 9a and 10a; Methods). Na1 is located approximately 10 Å from the cytoplasmic surface between TM4b and TM5, but also interacts with the carboxylate moiety of Glu 260 on TM9a (Fig. 3b and Supplementary Fig. 10a). The Na2 site is located 8 Å from Na1, near the centre at the crossover points of TM4a with TM4b and TM9a with TM9b. Four backbone carbonyl oxygen atoms coordinate Na2, including Glu 260 on TM9a and the side chains of Gln 264 on TM9a and Gln 77 on TM3. The residues for which the side chains interact with the two sodium ions are completely conserved in ASBT and NTCP (Supplementary Figs 5 and 8). The glutamate residue equivalent to Glu 260 is essential for activity in ASBT and NTCP^{13,22}. In ASBT_{NM}, its replacement with alanine significantly affects transport, as does the mutation of Gln 77 to alanine (Fig. 1d and Supplementary Fig. 2a). Thus, it seems that both sodium ions are required for efficient transport. Mechanistically, it is almost certainly necessary for sodium to be present at the Na2 site to neutralize the partial negative dipole of TM9a and, by doing so, stabilize the interaction with TM4a. Neutralization of the helix dipoles seems a conserved feature for this fold. In NhaA, the corresponding transmembrane helix is thought to be neutralized by the positive charge of Lys 300, which is essential for transport^{3,23}.

The substrate-binding cavity is open to the cytoplasm and is approximately 6 Å × 12 Å × 14 Å in size, with a solvent-accessible volume of 550 Å³ (Fig. 3a; Methods). Because the N-terminal half of TM1 is markedly bent outwards, it is more open on one side. The

cavity is much bigger than taurocholate, perhaps reflecting the large variety of compounds that are recognized by ASBT^{9,12,16} (Fig. 3a, c). It is predominantly hydrophobic, but near the bottom there are a number of polar residues and water molecules (Fig. 3c and Supplementary Fig. 10b). As judged from high B-factors, taurocholate seems to be weakly bound (Supplementary Table 2 and Supplementary Fig. 9b). Consistent with this observation, there is only one direct hydrogen bond between ASBT_{NM} and taurocholate, from Asn 295 on TM10 to the 7 α hydroxyl group. The mutation of Asn 295 to alanine causes a dramatic reduction in taurocholate transport (Fig. 1d and Supplementary Fig. 2a). Water molecules bridge the 7 α hydroxyl with His 294 and the 3 α hydroxyl with Asn 265, which is located at the crossover region of TM9. Thr 112 is also in the vicinity of the 3 α group but cannot be unambiguously placed. The 12 α hydroxyl group does not have any apparent hydrogen-bonding partner. The taurine moiety binds between TM1 and TM10. Interaction of the taurocholate with residues in TM10 is in agreement with biochemical data, which indicate that the last helix in ASBT has a dominant role in the translocation process²⁴. The location of Asn 265 between the TM4b and TM9b dipoles suggests that it may have a role in the mechanism. The importance of this residue has been inferred from mutagenesis studies on NTCP²². If Asn 265 is replaced by alanine in ASBT_{NM}, transporter activity is reduced by ~80% (Fig. 1d and Supplementary Fig. 2a). Although there are clear similarities between the binding sites in ASBT_{NM} and ASBT, there are also sequence differences (Supplementary Fig. 5). Such differences may affect substrate specificity.

For transport to take place, the protein must switch between outward- and inward-facing states²⁵. The architecture of ASBT_{NM} provides a clue to understanding how this might occur. The sodium ions are located in the core domain, close to the crossover points of the

discontinuous helices and occluded from the bulk solvent. In NhaA, sodium binding causes a rearrangement of these helices^{26,27}. In ASBT_{NM}, similar rearrangements in the core domain are therefore likely. Because NhaA translocates only protons and sodium ions²⁶ these helix movements might be sufficient for transport. However, ASBT_{NM} transports much larger substrates, and structural movements in more than the core domain are therefore needed. For the sodium-coupled transporter LeuT, the internal asymmetry of the repeating motifs has been used to predict global movements from a single structure²⁸, and these movements have been substantiated by crystallographic studies²⁹. In an analogous manner to LeuT, an outward-facing model of ASBT_{NM} was generated by superimposing TM1 to TM5 on TM6 to TM10, and vice versa (Fig. 4a; Methods). Comparing the inward-facing ASBT_{NM} structure with the outward-facing model, the largest difference is the position of the panel domain relative to the core domain (Fig. 4a, c). A route through the protein between these domains is in agreement with experimental data, which suggest that the final helix of ASBT and TM9 (IX) of NhaA line the transport pathway^{3,24,26,30}. Notably, the NhaA domain equivalent to the panel domain is located between that of the outward-facing and inward-facing ASBT_{NM} states (Fig. 4b). This may be because NhaA translocates a much smaller substrate, or it could represent another conformation of the transporter, probably an occluded state.

We propose that sodium binding controls the conformation of the core domain of ASBT_{NM}, which in turn drives the movement of the panel domain. This large conformational change of the panel domain relative to the core domain is required to alter the accessibility of the substrate-binding pocket. The ASBT_{NM} structure should aid the design of new inhibitors against ASBT with the goal of treating hypercholesterolaemia.

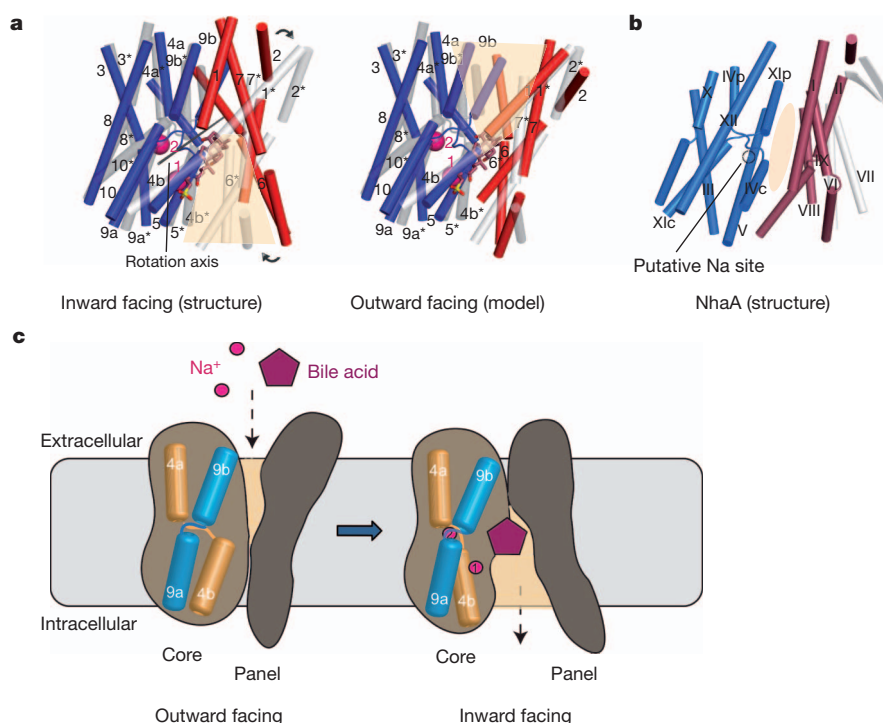


Figure 4 | Putative mechanism for ASBT_{NM} transport. **a**, Superposition of the ASBT_{NM} structure (red, panel domain; blue, core domain) and the outward-facing model as described in the text (light grey). The superposition has been optimized on the core domains. Loops have been removed for clarity. In the right-hand image, the panel domain of the model has been rotated by 25° relative to the core domain, around the axis shown in the left-hand image, to superimpose the panel domains. Significant kinks in the helices are represented as breaks. The area of the cavity is depicted by a tan-coloured trapezoid.

Transmembrane helices are numbered from 1 to 10; those coloured grey and labelled with an asterisk represent the helices in the outward-facing model. **b**, NhaA shown in the same view as ASBT_{NM} in **a**. The core domain is shown in light blue and the panel domain is shown in maroon. The two additional transmembrane helices and β -strands that are not present in ASBT_{NM} are shown in grey. The position where sodium is thought to bind³ is shown with a black ring. **c**, Proposed transport mechanism, illustrating the movement of the panel domain relative to the core domain to transport sodium and bile acid.

METHODS SUMMARY

ASBT_{NM} was cloned into a cleavable green fluorescent protein (GFP)/His₈ fusion vector, pWaldoGFP¹⁰. The fusion protein was expressed in *E. coli*, solubilized in 1% dodecyl- β -D-maltopyranoside and purified to homogeneity. Before crystallization, untagged ASBT_{NM} was exchanged into 0.06% *n*-dodecyl-*N,N*-dimethylamine-*N*-oxide by size-exclusion chromatography. Crystals were grown in the presence of 10 mM taurocholate by the vapour diffusion method. Data were collected on beamlines I02 and I03 at the Diamond Light Source, UK, dehydration of the crystals being necessary to collect high-resolution data. The protein was derivatized by short-soaking a surface-engineered cysteine mutant (ASBT_{NM-1}) with 1 mM mercury acetate. The structure of ASBT_{NM-1} was solved by single-wavelength anomalous dispersion and subsequently refined against data collected from ASBT_{NM} at a resolution of 2.2 Å. The cell-based bile acid uptake assay for ASBT_{NM} was modified from that previously described⁶.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 March; accepted 15 August 2011.

Published online 5 October 2011.

- Lewis, M. C., Brieady, L. E. & Root, C. Effects of 2164U90 on ileal bile acid absorption and serum cholesterol in rats and mice. *J. Lipid Res.* **36**, 1098–1105 (1995).
- Bhat, B. G. *et al.* Inhibition of ileal bile acid transport and reduced atherosclerosis in apoE^{-/-} mice by SC-435. *J. Lipid Res.* **44**, 1614–1621 (2003).
- Hunte, C. *et al.* Structure of a Na⁺/H⁺ antiporter and insights into mechanism of action and regulation by pH. *Nature* **435**, 1197–1202 (2005).
- Hallén, S., Björquist, A., Ostlund-Lindqvist, A. M. & Sachs, G. Identification of a region of the ileal-type sodium/bile acid cotransporter interacting with a competitive bile acid transport inhibitor. *Biochemistry* **41**, 14916–14924 (2002).
- Hagenbuch, B. & Dawson, P. The sodium bile salt cotransport family SLC10. *Pflügers Arch.* **447**, 566–570 (2004).
- Wong, M. H., Oelkers, P., Craddock, A. L. & Dawson, P. A. Expression cloning and characterization of the hamster ileal sodium-dependent bile acid transporter. *J. Biol. Chem.* **269**, 1340–1347 (1994).
- Weinman, S. A., Carruth, M. W. & Dawson, P. A. Bile acid uptake via the human apical sodium-bile acid cotransporter is electrogenic. *J. Biol. Chem.* **273**, 34691–34695 (1998).
- Oelkers, P., Kirby, L. C., Heubi, J. E. & Dawson, P. A. Primary bile acid malabsorption caused by mutations in the ileal sodium-dependent bile acid transporter gene (SLC10A2). *J. Clin. Invest.* **99**, 1880–1887 (1997).
- Kramer, W. & Wess, G. Bile acid transport systems as pharmaceutical targets. *Eur. J. Clin. Invest.* **26**, 715–732 (1996).
- Drew, D., Lerch, M., Kunji, E., Slotboom, D. J. & de Gier, J. W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nature Methods* **3**, 303–313 (2006).
- Sonoda, Y. *et al.* Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. *Structure* **19**, 17–25 (2011).
- Geyer, J., Wilke, T. & Petzinger, E. The solute carrier family SLC10: more than a family of bile acid transporters regarding function and phylogenetic relationships. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **372**, 413–431 (2006).
- Sun, A. Q., Balasubramanian, N., Chen, H., Shahid, M. & Suchy, F. J. Identification of functionally relevant residues of the rat ileal apical sodium-dependent bile acid cotransporter. *J. Biol. Chem.* **281**, 16410–16418 (2006).
- Chignard, N. *et al.* Bile acid transport and regulating functions in the human biliary epithelium. *Hepatology* **33**, 496–503 (2001).
- Craddock, A. L. *et al.* Expression and transport properties of the human ileal and renal sodium-dependent bile acid transporter. *Am. J. Physiol.* **274**, G157–G169 (1998).
- Zheng, X., Ekins, S., Raufman, J. P. & Polli, J. E. Computational models for drug inhibition of the human apical sodium-dependent bile acid transporter. *Mol. Pharm.* **6**, 1591–1603 (2009).
- Banerjee, A. & Swaan, P. W. Membrane topology of human ASBT (SLC10A2) determined by dual label epitope insertion scanning mutagenesis. New evidence for seven transmembrane domains. *Biochemistry* **45**, 943–953 (2006).
- Hallén, S., Branden, M., Dawson, P. A. & Sachs, G. Membrane insertion scanning of the human ileal sodium/bile acid co-transporter. *Biochemistry* **38**, 11379–11388 (1999).
- Scrapanti, E. & Hunte, C. Discontinuous membrane helices in transport proteins and their correlation with function. *J. Struct. Biol.* **159**, 261–267 (2007).
- Boudker, O. & Verdon, G. Structural perspectives on secondary active transporters. *Trends Pharmacol. Sci.* **31**, 418–426 (2010).
- Hagenbuch, B. & Meier, P. J. Sinusoidal (basolateral) bile salt uptake systems of hepatocytes. *Semin. Liver Dis.* **16**, 129–136 (1996).
- Zahner, D., Eckhardt, U. & Petzinger, E. Transport of taurocholate by mutants of negatively charged amino acids, cysteines, and threonines of the rat liver sodium-dependent taurocholate cotransporting polypeptide Ntcp. *Eur. J. Biochem.* **270**, 1117–1127 (2003).
- Olkhova, E., Hunte, C., Scrapanti, E., Padan, E. & Michel, H. Multiconformation continuum electrostatics analysis of the NhaA Na⁺/H⁺ antiporter of *Escherichia coli* with functional implications. *Proc. Natl Acad. Sci. USA* **103**, 2629–2634 (2006).
- Hussainzada, N., Banerjee, A. & Swaan, P. W. Transmembrane domain VII of the human apical sodium-dependent bile acid transporter ASBT (SLC10A2) lines the substrate translocation pathway. *Mol. Pharmacol.* **70**, 1565–1574 (2006).
- Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).
- Padan, E. The enlightening encounter between structure and function in the NhaA Na⁺-H⁺ antiporter. *Trends Biochem. Sci.* **33**, 435–443 (2008).
- Appel, M., Hizlan, D., Vinothkumar, K. R., Ziegler, C. & Kuhlbrandt, W. Conformations of NhaA, the Na⁺/H⁺ exchanger from *Escherichia coli*, in the pH-activated and ion-translocating states. *J. Mol. Biol.* **388**, 659–672 (2009).
- Forrest, L. R. *et al.* Mechanism for alternating access in neurotransmitter transporters. *Proc. Natl Acad. Sci. USA* **105**, 10338–10343 (2008).
- Shimamura, T. *et al.* Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* **328**, 470–473 (2010).
- Tzuber, T., Rimon, A. & Padan, E. Structure-based functional study reveals multiple roles of transmembrane segment IX and loop VIII-IX in NhaA Na⁺/H⁺ antiporter of *Escherichia coli* at physiological pH. *J. Biol. Chem.* **283**, 15975–15987 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to C. Lee and Y. Sekiguchi for assistance with cloning and expression screening of ASBT_{NM} mutants, and to S. van de Graaf for donating fluorescently labelled bile acid, which was used in the initial functional characterization of ASBT_{NM}. Data were collected at the European Synchrotron Radiation Facility, France, and the Diamond Light Source, UK, with assistance from beamline scientists; in particular, we would like to thank J. Sanchez-Weatherby for help with the HCl. We are also grateful to K. Beis and G. von Heijne for reading the manuscript. This work was funded by the Medical Research Council (MRC_G0900990/91997), to A.D.C. and D.D.), the European Union (EMeP grant LSHG-CT-2004-504601, to S.I.) and the Biotechnology and Biological Sciences Research Council (BB/G023425/1, to S.I.). Part of this work was also supported by a grant from the Targeted Proteins Research Program of MEXT, Japan, and the ERATO Iwata Human Receptor Crystallography Project, Japan Science and Technology Agency. The authors are grateful for the use of the Membrane Protein Laboratory funded by the Wellcome Trust (WT089809) at the Diamond Light Source. D.D. acknowledges personal support from The Royal Society through the University Research Fellow scheme.

Author Contributions N.-J.H., S.I., A.D.C. and D.D. contributed to the design of the project. N.-J.H. and D.D. screened homologues, expressed and purified the protein, and carried out functional characterization. N.-J.H., S.I., A.D.C. and D.D. were involved in crystallographic experiments and analysis of data. A.C. and D.D. were responsible for overall project management and wrote the manuscript with assistance from N.-J.H. and S.I.

Author Information The coordinates and the structure factors for ASBT_{NM} and ASBT_{NM-1} have been deposited in the Protein Data Bank under accession numbers 3ZUY and 3ZUX, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.D. (d.drew@imperial.ac.uk), A.D.C. (a.cameron@imperial.ac.uk) or S.I. (s.iwata@imperial.ac.uk).

METHODS

ASBT_{NM} sequence. MNILSKISSFIGKTFSLWAALFAAAFFAPDTFKWAGPY IPWLLGIIMFGMGLTLKPSDFDILFKHPKVVIIGVIAQFAIMPATAWLLSKLL NPALAEIAGVILVGCCPGGTASNVMTYLARGNVALSVAVTSVSTLSPILLTP AIFMLLAGEMLEIQAAGMLMSIVKMLVLLPIVLGLIVHKVLGSKTEKLTALP LVSVAIVLIIGAVVGASKGKIMESGLLFAVVVLHNGIGYLLGFFAAKWTG LPYDAQKTLTIEVGMQNSGLAALAAAHFAAPVAVPGALFSVWHNISG SLLATYWAAKAGKHKKPGSENLYFQ

ASBT_{NM-1} sequence for structure solution. MVAASMNILSKISSFIGKTFSLW AALFAAAFFAPDTFKWAGPYIPWLLGIIMFGMGLTLKPSDFDILFKHPKV VIIGVIAQFAIMPATAWCLSKLLNPALAEIAGVILVGCCPGGTASNVMTYL ARGNVALSVAVTSVSTLTPAIFMLLAGEMLEIQAAGMLMSIVKMLV LPIVLGLIVHKVLGSKTEKLTALP LVSVAIVLIIGAVVGASKGKIMESGLL IFAVVVLHNGIGYLLGFFAAKWTG LPYDAQKALTIEVGMQNSGLAALAA AHFAAPVAVPGALFSVWHNISG SLLATYWAAKAGKHKKPLDRAGSEN LYFQ

Expression screening, mutagenesis and protein purification. Bacterial ASBT homologues were cloned as GFP–His₈ fusions into the vector pWaldoGFP¹⁰, as fluorescence from a C-terminal GFP fusion is a reliable reporter of membrane-integrated expression³¹. Fusions were overexpressed in *E. coli* C43(DE3) cells³² by the addition of 0.4 mM IPTG at A_{600nm} 0.4. The temperature was decreased to 25 °C for overnight induction. The monodispersity of expressed fusions were screened in crude DDM-, decyl-β-D-maltopyranoside-, nonyl-β-D-maltopyranoside-, LDAO- or dodecyl nonaethylene glycol ether (C₁₂E₉)-solubilized membranes by fluorescence-detection size-exclusion chromatography³³ (FSEC). The ASBT_{NM} homologue from *N. meningitidis* (MC58) was selected for structural studies on the basis of the amount of protein produced, as judged by whole-cell³¹ and in-gel fluorescence¹⁰ and the quality of the FSEC trace in different detergents. Site-directed mutants of ASBT_{NM} were generated by PCR (Quickchange, Agilent Technologies).

Wild-type ASBT_{NM} and mutants were purified essentially as previously described³⁴. In brief, membranes were isolated from 10-l *E. coli* cultures and solubilized in 1% DDM for 2 h in buffer containing ×1 PBS, 150 mM NaCl and 10 mM imidazole. The suspension was cleared by ultracentrifugation at 120,000g for 1 h. The sample was mixed with 1 ml of Ni-NTA Superflow resin (QIAGEN) per 1 mg of GFP–His₈ and incubated for 2 h at 4 °C. Slurry was loaded onto a glass Econo-Column (Bio-Rad) and washed in ×1 PBS buffer containing 0.1% DDM, 150 mM NaCl and 20 mM imidazole for 20 column volumes. Bound material was washed for a further 20 column volumes in the same buffer containing 50 mM imidazole. The ASBT_{NM}–GFP–His₈ fusion was eluted in two column volumes of the same buffer containing 250 mM imidazole. The eluted protein was dialysed overnight in the presence of stoichiometric amounts of His₆-tagged tobacco etch virus protease in 3 l of buffer containing 20 mM Tris-HCl, pH 7.5, 150 mM NaCl and 0.03% DDM. Dialysed sample was passed through a 5-ml Ni-NTA His-Trap column (GE Healthcare), and the flow-through containing ASBT_{NM} was collected. Protein was concentrated to 10 mg ml⁻¹ using concentrators with a relative molecular mass cut-off of 100K, and was loaded onto a Superdex 200 10/300 gel filtration column (GE Healthcare) equilibrated in 20 mM Tris-HCl, pH 7.5, 150 mM NaCl and 0.06% LDAO. The choice of the detergent LDAO was considered suitable for crystallization by comparing FSEC³³ and stability data³⁵ for ASBT_{NM} with membrane proteins known to crystallize in this detergent¹¹. The protein peak was collected and concentrated to 20 mg ml⁻¹ for crystallization.

Transport time course. *E. coli* cells harbouring wild-type ASBT_{NM}–GFP–His₈ were collected and resuspended in uptake buffer consisting of 1 mM CaCl₂, 1 mM MgCl₂, 10 mM Tris-HCl, pH 7.5, and either 137 mM NaCl (Na⁺-containing buffer) or 137 mM choline chloride (Na⁺-low buffer). Cells were incubated at 37 °C with uptake buffer containing 4 μM taurocholate supplemented with 0.16 μM [2,4-³H]-taurocholate (30 Ci mmol⁻¹; American Radiolabelled Chemicals) for the indicated time intervals. Transport was terminated by the addition of ice-cold buffer containing 1 mM CaCl₂, 1 mM MgCl₂, 10 mM Tris-HCl, pH 7.5, 137 mM NaCl or 137 mM ChCl, and 1 mM taurocholate, and was followed immediately by centrifugation at 20,500g for 60 s. Cell pellets were washed several times in an equal volume of termination buffer and resuspended in 200 μl of the same buffer. The radioactivity corresponding to the internalized substrate was measured by scintillation counting. Each experiment was performed in triplicate. Nonspecific uptake was assessed by repeating the time course in triplicate for cells transformed with the same vector but expressing the sodium/proton antiporter fusion NhaA–GFP–His₈. In all experiments, ASBT_{NM} expression was calculated on the basis of GFP fluorescence measured at 510 nm (excitation wavelength, 488 nm) using a 96-well spectrofluorometer³¹. In-gel fluorescence and FSEC data of DDM-solubilized whole cells of wild-type ASBT_{NM} and mutants were carried out essentially as described previously^{10,33}.

Transport kinetics. The accumulation of taurocholate was linear within the first 120 s. For kinetic characterization, the initial velocity of taurocholate uptake at

37 °C was measured after 120 s at the indicated increasing substrate concentrations. The radioactivity corresponding to the internalized substrate was measured by scintillation counting. Nonspecific uptake was measured by repeating the transport kinetics for cells transformed with the same vector, but expressing the sodium/proton antiporter fusion NhaA–GFP–His₈. Specific ASBT_{NM} uptake was calculated by the subtraction of nonspecific uptake from total uptake. Each experiment was performed in triplicate. The data were fitted to the Michaelis–Menten equation by nonlinear regression using the GRAPHPAD PRISM software.

Activity of ASBT_{NM} mutants. *E. coli* cells harbouring ASBT_{NM}–GFP–His₈ mutants were resuspended in uptake buffer containing 4 μM taurocholate supplemented with 0.16 μM [2,4-³H]-taurocholate (30 Ci mmol⁻¹; American Radiolabelled Chemicals) for 5 min at 37 °C. The radioactivity corresponding to the internalized substrate was measured by scintillation counting. For each mutant, the uptake values were corrected for background by subtracting values from parallel assays carried out in the absence of sodium. Activities were plotted as percentages of the wild-type transport activity calculated in the same way. Each experiment was performed in triplicate.

Substrate specificity. The whole-cell [³H]-taurocholate uptake assay was carried out similarly to that described for ASBT_{NM} mutants, except that 150 μM of either taurocholate (Sigma), cyclosporin A (Sigma), bromosulphophthalein (Sigma) or fluvastatin (Cayman Europe) was added to the uptake buffer.

Crystallization and preliminary screening. Crystals were grown at 20 °C using the vapour diffusion method. Taurocholic acid (Sigma) was added to the protein solution to a final concentration of 10 mM. The protein was then mixed 1:1 with reservoir solution containing 50 mM sodium citrate, pH 4.5, 70 mM NaCl and 22–24% PEG 400. Crystals appeared overnight and reached a maximum size after 3–4 days. The crystals were frozen in liquid nitrogen and screened using synchrotron radiation at the European Synchrotron Radiation Facility and Diamond Light Source. Crystals are tetragonal with cell dimensions of approximately 75 Å × 75 Å × 180 Å. The best of these crystals diffract to around 2.8–3.5 Å; however, with dehydration the diffraction increases to ~2 Å.

Structure determination of a cysteine mutant of ASBT_{NM-1}. As initial attempts at making heavy-atom derivatives with mercury compounds failed, Leu 87 was modified to cysteine (construct ASBT_{NM-1}). The ASBT_{NM-1} protein crystallized similarly to the wild-type protein. Mercury-derivatized crystals were obtained from this mutant by incubating for 1 h with 1 mM mercury acetate before crystallization. A single mercury-derivatized crystal of ASBT_{NM-1} was used to solve the structure by single-wavelength anomalous dispersion. The crystal was frozen in liquid nitrogen and then re-annealed before data collection by leaving it in air for approximately 3 min. The re-annealing resulted in shrinkage of the unit cell and an increase in the resolution to 2.2 Å. Data were collected at the mercury edge (1.0660 Å) on beamline I03 at the Diamond Light Source.

Data were initially processed to 2.5 Å by the XIA2³⁶ pipeline to XDS³⁷ set up on the beamline, with further processing using the CCP4 suite of programs³⁸. The space group was determined to be P₄2₁2, with one molecule in the asymmetric unit. An anomalous difference Patterson map showed clear peaks associated with one bound heavy atom. The heavy-atom coordinates were determined using RSPS³⁹. Its position was refined and phases were calculated using SHARP⁴⁰ with solvent flattening in SOLOMON⁴¹. The resulting phases were input to the automatic structure building implemented in PHENIX⁴². This resulted in a model that was reasonably complete. Modification and further building of the structure was carried out in O⁴³ and COOT⁴⁴. At this point, the data were reprocessed using MOSFLM⁴⁵, extending the resolution to 2.2 Å as judged from the scaling statistics (Supplementary Table 1) and the features in the resulting maps.

Structural refinement was performed in BUSTER⁴⁶ using individual isotropic B-factor refinement and TLS⁴⁷. The complete protein was chosen as a single TLS group because no significant drop in the *R*_{free} value was observed when splitting the protein into multiple groups.

Two ions were identified in the core of the protein. The residues coordinating these ions and the associated distances are consistent with their being sodium ions⁴⁸. As an additional verification, the putative sodium ions were changed to water molecules and run through the program WASP⁴⁹, which uses valence calculations to identify possible metal ions. After replacing all solvent and ions by water molecules as required by the program, only the two solvent molecules originally assigned as sodium ions were flagged as likely sodium ions. After all residues had been modelled, clear electron density remained in the cavity of the protein. This density was enhanced in a simulated-annealing omit map calculated in PHENIX⁴². The taurocholate structure, downloaded from the Cambridge Structural Database (accession code, KORZUM), clearly fitted the density with the cholate head group positioned at the bottom of the cavity (Supplementary Fig. 9b). A further taurocholate was observed in the crystal interface. The final model has an *R*-factor of 19.7% and a corresponding *R*_{free} value of 22.9%, and contains all protein residues from 2 to 309, two sodium ions, one mercury atom,

two taurocholate molecules, 37 water molecules, five LDAO molecules and two truncated phospholipids (phosphatidylethanolamine). The final refinement statistics of this model, which was used to solve the wild-type protein, are summarized in Supplementary Table 2.

Structure determination and refinement of ASBT_{NM}. Because the re-annealing of the ASBT_{NM-1} construct in air was not reproducible, dehydration was attempted using the humidity controller HC1 device⁵⁰ mounted on beamline I02 at the Diamond Light Source. By placing the crystal into an air stream at 45% relative humidity for 5 min before freezing it, crystals were found to reproducibly diffract to ~ 2.0 Å. Data were collected from a single crystal of ASBT_{NM} on beamline I02 at the Diamond Light Source. The data were processed in XDS³⁷ using the XIA2 pipeline³⁶ and scaled at a resolution of 2.2 Å (Supplementary Table 1). The structure was refined, as above, starting from the final model of the ASBT_{NM-1} construct, minus all non-protein residues. No appreciable differences were observed in the wild-type and mercury-derivatized structures. Taurocholate and detergent molecules were modelled in the same positions as for ASBT_{NM-1}. The final model has an *R*-factor of 21.2% and an *R*_{free} value of 24.4% (Supplementary Table 2).

Structural analysis. Superpositions were carried out in LSQMAN⁵¹. The superpositions were performed so that only C α pairs that were less than 3.8 Å apart were included in the calculation. The numbers quoted in the text regarding the topology-inverted repeats of ASBT_{NM} were calculated between pairs of C α atoms that were less than 10 Å apart. This was considered necessary to include atoms from both the V-motif and the core motif. In comparing ASBT_{NM} with NhaA (Protein Data Bank ID, 1ZCD), only pairs of atoms less than 5 Å apart after superposition were chosen, giving an r.m.s.d. of 2.9 Å for 202 out of a possible 308 pairs of C α atoms. The volume of the cavity was calculated in VOIDOO⁵² using a probe radius of 1.4 Å. Figures showing the structure were drawn using PYMOL⁵³ except those showing electron density, which were made using CCP4MG⁵⁴.

Outward-facing model. As in LeuT⁵⁵, in ASBT_{NM} the protein is made up of two five-transmembrane-helix repeats that when superimposed show a small rotation of two transmembrane helices with respect to the other three (Supplementary Fig. 4). For LeuT, it was shown that by swapping the conformations of the N- and C-terminal topology-inverted repeats the structure can be changed from an outward-facing state to an inward-facing state²⁸. In ASBT_{NM}, the lengths of the two topology-inverted repeats are very similar. To create an outward-facing backbone model of ASBT_{NM}, in an analogous manner to that carried out for LeuT, TM1 to TM5 were superposed on TM6 to TM10, and vice versa.

31. Drew, D. E., von Heijne, G., Nordlund, P. & de Gier, J. W. Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. *FEBS Lett.* **507**, 220–224 (2001).
32. Miroux, B. & Walker, J. E. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298 (1996).

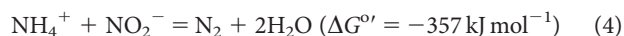
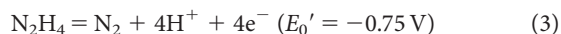
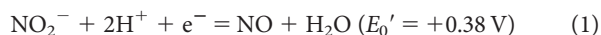
33. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
34. Drew, D. *et al.* GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nature Protocols* **3**, 784–798 (2008).
35. Alexandrov, A. I., Mileni, M., Chien, E. Y., Hanson, M. A. & Stevens, R. C. Microscale fluorescent thermal stability assay for membrane proteins. *Structure* **16**, 351–359 (2008).
36. Winter, G. Xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
37. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
38. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
39. Knight, S. D. RSPS version 4.0: a semi-interactive vector-search program for solving heavy-atom derivatives. *Acta Crystallogr. D* **56**, 42–47 (2000).
40. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
41. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
42. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
43. Jones, T. A. & Kjeldgaard, M. Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208 (1997).
44. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
45. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **26** (1992).
46. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D* **60**, 2210–2221 (2004).
47. Winn, M. D., Isupov, M. N. & Murshudov, G. N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133 (2001).
48. Harding, M. M. Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallogr. D* **58**, 872–874 (2002).
49. Nayal, M. & Di Cera, E. Valence screening of water in protein crystals reveals potential Na⁺ binding sites. *J. Mol. Biol.* **256**, 228–234 (1996).
50. Sanchez-Weatherby, J. *et al.* Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallogr. D* **65**, 1237–1246 (2009).
51. Kleywegt, G. J. & Jones, T. A. A super position. *ESF/CCP4 Newslett.* **31**, 9–14 (1994).
52. Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D* **50**, 178–185 (1994).
53. DeLano, W. L. *PyMOL Molecular Viewer* (<http://www.pymol.org>) (2002).
54. Potterton, L. *et al.* Developments in the CCP4 molecular-graphics project. *Acta Crystallogr. D* **60**, 2288–2294 (2004).
55. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na⁺/Cl[−]-dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).

Molecular mechanism of anaerobic ammonium oxidation

Boran Kartal¹, Wouter J. Maalcke¹, Naomi M. de Almeida¹, Irina Cirpus¹, Jolein Gloerich², Wim Geerts¹, Huub J. M. Op den Camp¹, Harry R. Harhangi¹, Eva M. Janssen-Megens³, Kees-Jan Francoijs³, Hendrik G. Stunnenberg³, Jan T. Keltjens¹, Mike S. M. Jetten^{1,4} & Marc Strous^{1,5}

Two distinct microbial processes, denitrification and anaerobic ammonium oxidation (anammox), are responsible for the release of fixed nitrogen as dinitrogen gas (N₂) to the atmosphere^{1–4}. Denitrification has been studied for over 100 years and its intermediates and enzymes are well known⁵. Even though anammox is a key biogeochemical process of equal importance, its molecular mechanism is unknown, but it was proposed to proceed through hydrazine (N₂H₄)^{6,7}. Here we show that N₂H₄ is produced from the anammox substrates ammonium and nitrite and that nitric oxide (NO) is the direct precursor of N₂H₄. We resolved the genes and proteins central to anammox metabolism and purified the key enzymes that catalyse N₂H₄ synthesis and its oxidation to N₂. These results present a new biochemical reaction forging an N–N bond and fill a lacuna in our understanding of the biochemical synthesis of the N₂ in the atmosphere. Furthermore, they reinforce the role of nitric oxide in the evolution of the nitrogen cycle.

Ammonium is difficult to activate in the absence of molecular oxygen. Therefore, how anammox bacteria are able to oxidize ammonium coupled to the reduction of nitrite and forge an N–N bond to make N₂ has been an intriguing question for a long time. Based on the *in silico* analysis of the genome assembly of the anammox bacterium *Kuenenia stuttgartiensis*, a set of three redox reactions (equations (1)–(3)) involving N₂H₄ and nitric oxide (NO) was proposed⁶ to explain the overall anammox stoichiometry (equation (4)):



The role of N₂H₄ in anammox catabolism was originally proposed based on the observation that the compound transiently accumulated when anammox bacteria were incubated with millimolar quantities of hydroxylamine^{7,8}. However, the turnover of neither N₂H₄, hydroxylamine nor NO was demonstrated to start from the actual substrates ammonium and nitrite; thus it remained unclear whether the observed reaction was an integral part of the anammox pathway or a side reaction.

In the present study, we resolved the anammox pathway and its enzymes by a combination of complementary approaches (Fig. 1). *K. stuttgartiensis* was enriched and grown as suspended cells in a membrane bioreactor^{9,10}. Fluorescence *in situ* hybridization (FISH) showed that *K. stuttgartiensis* made up more than 95% of the population. Transcription was shown for more than 97% of all genes after random hexamer-primed reverse transcription of extracted RNA, sequencing and mapping of 5.6 million 32-nucleotide reads on an Illumina Genome Analyser (metatranscriptome accession number

GSE15408). Expression of 1010 proteins was demonstrated by meta-proteomics¹¹ (peptidome accession number PSE111). Further, inhibitor and isotope labelling studies were performed and the activity of enzyme complexes was demonstrated after their purification by liquid chromatography.

Transcriptomics and proteomics indicated that *K. stuttgartiensis* expressed *cd1* nitrite:nitric oxide reductase (NirS, kuste4136, 9% of predicted peptides detected (p.p.d.) and 6.3-fold messenger RNA (mRNA) coverage) with the potential ability to reduce nitrite to NO. This possibility was investigated by incubating cell suspensions of *K. stuttgartiensis* with ammonium, nitrite (2 mM each) and 100 μM NO scavenger PTIO (2-phenyl-4,4,5,5-tetramethylimidazole-1-oxyl-3-oxide)¹². When PTIO was introduced at the start of the incubation or when it was added to active cells, anammox activity was inhibited (Fig. 2a). Further, the cells were incubated with ammonium and nitrite (2 mM each) in the presence of DAF2-DA (10 μM) that reacts with NO to form a fluorescent product^{13,14}. Sampled *K. stuttgartiensis* cells displayed the characteristic green fluorescence indicating NO production (Fig. 2b and Supplementary Fig. 1). In control experiments without nitrite or with added PTIO, there was no detectable fluorescent signal. It should be noted that both PTIO and DAF2-DA might have a wider reaction spectrum than NO and might

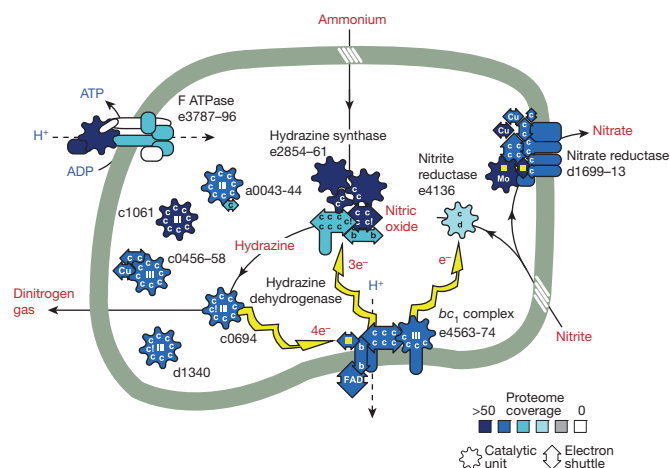


Figure 1 | Biochemical pathway and enzymatic machinery of *K. stuttgartiensis*. The anammoxosome, an intracytoplasmic compartment bounded by a membrane (grey line), is the locus of anammox catabolism. Identifiers of open reading frames and the degree to which the encoded respiratory protein complexes were detected in the proteome are indicated. Hydrazine synthase depicted in the centre of the figure is also loosely membrane associated. Yellow arrows, electron flow; yellow square, iron-sulphur clusters; b, haem b; c, haem c; cl, atypical haem c; d, haem d; Mo, molybdopterin. Cofactors and motifs were determined previously⁶.

¹Institute for Water and Wetland Research, Department of Microbiology, Radboud University Nijmegen, Heyendaalseweg 135, 6525AJ Nijmegen, The Netherlands. ²Nijmegen Centre for Mitochondrial Disorders, Nijmegen Proteomics Facility, Department of Laboratory Medicine, Radboud University Nijmegen Medical Centre, Geert Grooteplein 10, 6500HB Nijmegen, The Netherlands. ³Radboud University, Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Geert Grooteplein 28, 6525 GA Nijmegen, The Netherlands. ⁴Delft University of Technology, Department of Biotechnology, 2628 BC Delft, The Netherlands. ⁵Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany.

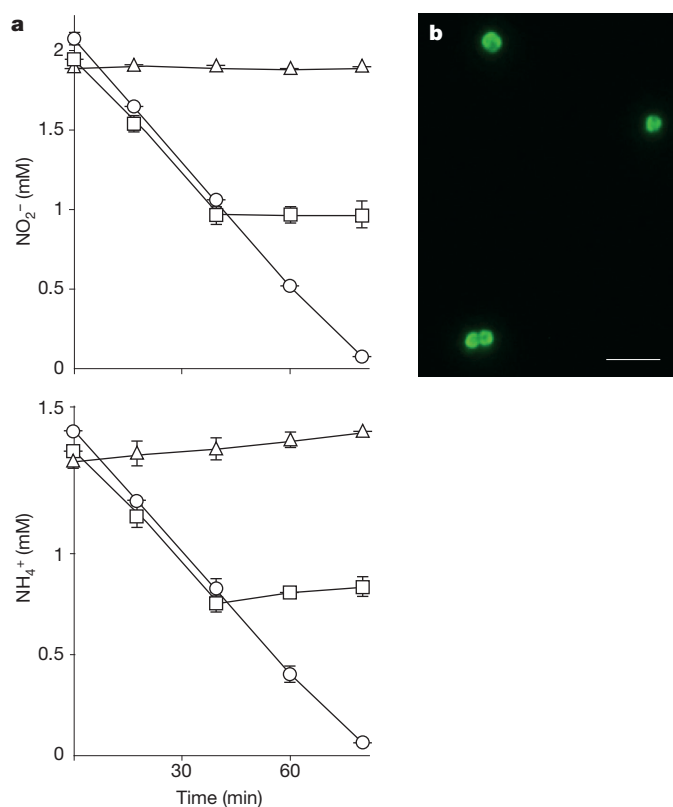
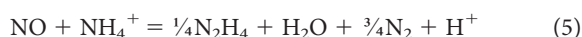


Figure 2 | Determination of nitric oxide (NO) as an intermediate. NO_2^- and NH_4^+ (2 mM each) conversion was inhibited by 100 μM PTIO (a). PTIO added at $t = 0$ (open triangle), PTIO added at 40 min (open square) and without PTIO (open circle), $n = 2$ (error bars, s.d.). (b) Epifluorescence image of (diaminofluorescein-2-diacetate) DAF2-DA derivative of NO formed during NH_4^+ and NO_2^- (2 mM each) conversion by anammox bacteria (scale bar, 10 μm).

possibly react with other nitrogen monoxides such as nitroxyl (HNO). However, unlike NO, HNO was not a suitable substrate for hydrazine synthase (see below).

Interestingly, when acetylene (15 μM) was added, the anammox reaction was inhibited. Acetylene inhibits aerobic ammonia monooxygenase, the ammonia-activating enzyme of aerobic ammonium oxidizers^{15–17}. Apparently, it also interfered with the ammonium-activating step of anammox cells (equation (2)). Importantly, acetylene inhibition resulted in an immediate accumulation of NO; hydroxylamine accumulation was not observed, consistent with the role of NO as the direct precursor for N_2H_4 .

The second step of the predicted anammox pathway would then be the reduction of NO and its simultaneous condensation with ammonium to produce N_2H_4 (equation (2)). Because the role of N_2H_4 in anammox catabolism was not established, we first demonstrated its *in vivo* turnover (Fig. 3a, b). To investigate whether N_2H_4 could be produced directly from NO, cell suspensions were incubated with NO (0.1 mM) and ammonium (2 mM). A transient accumulation of hydrazine (18 μM) was observed (Fig. 3c), albeit at a much lower concentration (200–500 μM) for incubations with hydroxylamine^{7,8}. This is consistent with equations (1)–(3) because the major part of the produced N_2H_4 would be oxidized to N_2 as expected from the overall reaction and NO could be supplied at much lower concentrations (equation (5)).



The anammox pathway is completed by the oxidation of N_2H_4 to N_2 (equation (3)). For a long time, N_2H_4 was known as an alternative

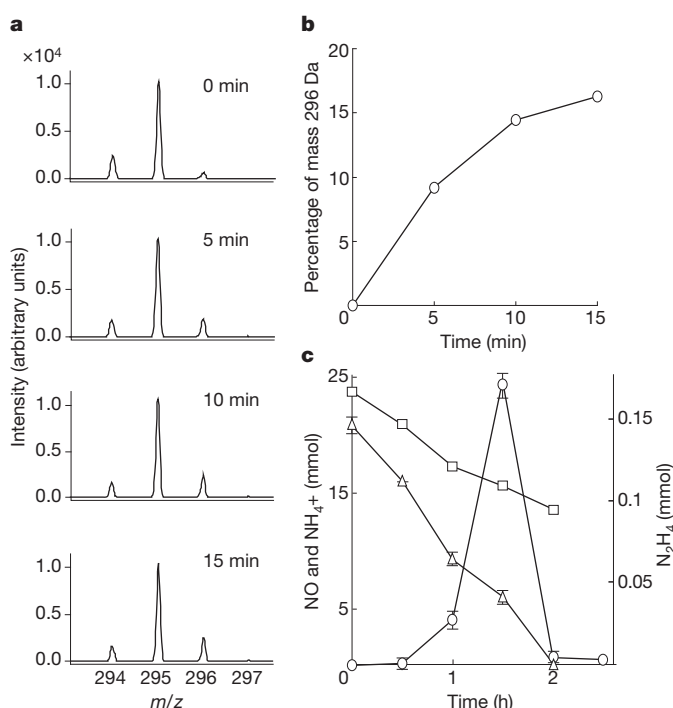


Figure 3 | Hydrazine turnover. *K. stuttgartiensis* cells were incubated with 2 mM $^{15}\text{NO}_2^-$ and $^{14}\text{NH}_4^+$ each in the presence of 2 mM $^{28}\text{N}_2\text{H}_4$. Under these conditions cells would only produce $^{29}\text{N}_2\text{H}_4$ and preferentially consume $^{28}\text{N}_2\text{H}_4$, leading to ^{15}N -label accumulation in the N_2H_4 pool. The 295 and 296 m/z masses correspond to derivatization products of $^{28}\text{N}_2\text{H}_4$ and $^{29}\text{N}_2\text{H}_4$ with *para*-dimethylaminobenzaldehyde³⁰ (a). The 294 m/z mass arises from the impurities of the matrix. Within 15 min, 16% of the N_2H_4 pool was labelled (b). Hydrazine (open circles) was produced by the cells incubated with 2 mM NH_4^+ (open triangles) and NO (0.1 mM) (open squares), $n = 2$ (error bars, s.d.) (c).

substrate for octahaem hydroxylamine oxidoreductases (HAOs), the enzymes that catalyse the conversion of hydroxylamine to nitrite in aerobic ammonium oxidizers^{18,19}. Strikingly, the *K. stuttgartiensis* genome encoded ten divergent paralogues of this enzyme, and six were detected at high levels in the transcriptome and proteome (mRNA up to 189-fold coverage, 27–58% p.p.d.; Supplementary Table 1). Six expressed paralogues belonged to the ‘type II’ hydrazine/hydroxylamine oxidoreductases (HZO/HAO)²⁰. Two related ‘type II’ HZO/HAO and one divergent octahaem cytochrome *c* were also detected at lower levels (4–15% p.p.d.) and one was not detected. By a two-step liquid chromatography procedure, we purified two highly expressed HZO/HAO-like proteins (kustc0694 and kustc1061). These enzymes appear to be closely related to two enzymes of unknown function isolated from an anammox enrichment culture KSU-1 (refs 21, 22). Both enzymes catalysed the four-electron oxidation of N_2H_4 to N_2 with cytochrome *c* as the artificial electron acceptor with different rates (2.5 and 0.4 $\mu\text{mol min}^{-1} \text{mg protein}^{-1}$, respectively). When they were incubated with $^{30}\text{N}_2\text{H}_4$ and cytochrome *c*, $^{30}\text{N}_2$ ($^{15}\text{N}^{15}\text{N}$) was produced stoichiometrically, in agreement with equation (3). Interestingly, Kustc1061 also oxidized hydroxylamine to NO (rather than nitrite) with a higher rate (6 $\mu\text{mol min}^{-1} \text{mg protein}^{-1}$). In contrast, kustc0694 did not catalyse this reaction and hydroxylamine and NO were powerful inhibitors of N_2H_4 oxidation, suggesting kustc0694 was the dedicated hydrazine dehydrogenase (HDH) in *K. stuttgartiensis*. Furthermore, the inhibition of kustc0694 explained the transient accumulation of N_2H_4 in the presence of hydroxylamine or NO.

Although no enzyme is known to convert NO and ammonium into N_2H_4 , two candidate gene clusters were previously identified potentially encoding an enzyme complex with this function (hydrazine synthase, HZS)⁶. One of these clusters (kuste2859–61) encoded the

most highly expressed proteins in the proteome (greater than 60% p.p.d., visible as three dominant spots on two-dimensional gels; Supplementary Table 1 and Supplementary Fig. 2a) and extremely abundant mRNAs in the transcriptome (greater than 50-fold coverage). The transcription of the other candidate cluster (kuste2474–83) was well below average (1.7-fold coverage) and expression was not detected by proteomics.

The kuste2859–61 proteins were purified from the cell-free extract of the *K. stuttgartiensis* as a complex that separated into three distinct bands on a denaturing polyacrylamide gel, corresponding to polypeptides encoded by three consecutive genes (kuste2859–2860–2861, Supplementary Fig. 2). Native polyacrylamide gel electrophoresis revealed that the complex was a multimer of approximately 240 kDa. Hydrazine synthesis activity of the complex was shown in a coupled assay with the kusc1061 HZO/HAO, using ^{15}N -ammonium (1 mM) and NO (0.9 mM) as substrates (Fig. 4). In the assay, kusc1061 would ‘pull’ the reaction by rapidly oxidizing the produced N_2H_4 to $^{29}\text{N}_2$ as the end product, while simultaneously ‘pushing’ the reaction by providing the electrons for N_2H_4 synthesis (equations (2) and 3). Kusc1061 alone did not catalyse the reaction, and N_2 production above background could not be measured in the absence of ammonium or NO. N_2 was not produced above background when hydroxylamine or nitroxyl (HNO) were provided as substrates with ammonium. The activity of N_2 formation in the coupled assay was $20 \text{ nmol h}^{-1} \text{ mg protein}^{-1}$, lower than the activity of whole cells with ammonium and nitrite (approximately $1800 \text{ nmol h}^{-1} \text{ mg protein}^{-1}$). The cell-free extracts were unable to form N_2 from ammonium and nitrite, but could from NO and ammonium under the same experimental conditions, at six-fold lower rate than the purified HZS ($3.4 \text{ nmol h}^{-1} \text{ mg protein}^{-1}$). The decrease in activity upon mere cell disruption was most probably due to the disruption of a tightly coupled multi-component system with hydrazine synthesis as the rate-limiting step.

Interestingly, the kuste2859–61 complex was capable of N_2 formation from ammonium and NO on its own (Fig. 4). The purified enzyme oxidized N_2H_4 to N_2 with a specific activity of $34 \text{ nmol min}^{-1} \text{ mg protein}^{-1}$, resulting in an overall disproportionation reaction (equation (5)). Considering that N_2H_4 is the energy source in anammox

metabolism, N_2 formation by HZS would be unproductive. Consequently, we may speculate that the anammox bacterium harbours backup systems that efficiently trap hydrazine and that keep (nitrogenous) inhibitory compounds, like NO and hydroxylamine, at low concentrations, which would partly explain the redundancy of HAO/HZO-like proteins in the organism. Our experiments showed that HZS and HDH were necessary and sufficient to make N_2 from the substrate ammonium and the intermediate NO.

Taken together, anammox catabolism and energy for growth must be conserved from three reactions (equations (1)–(3)). It is hypothesized that anammox bacteria synthesize ATP through a membrane-bound ATP synthase complex driven by proton-motive force (pmf) generated through catabolic reactions with the intermediary action of the quinol:cytochrome *c* oxidoreductase system (complex III, the bc_1 complex).

Intiguously, three gene clusters encoding bc_1 complexes and four encoding ATP synthases were present in the *K. stuttgartiensis* genome. Transcription and expression of one (kuste4569–74) of these gene clusters were detected at higher levels (26–33% p.p.d., 6- to 24-fold mRNA coverage) than the other two (0–19% p.p.d., 2- to 15-fold mRNA coverage). When *K. stuttgartiensis* cell suspensions were spiked with pentachlorophenol (10 μM), a structural analogue of quinol and a known inhibitor of the bc_1 complex, anammox activity was completely inhibited, indicating that the bc_1 complex was involved in energy conservation and its role in electron transport from N_2H_4 oxidation to nitrite reduction and hydrazine synthesis was not backed up by any other system. The expression of the four gene clusters encoding ATP synthase was even more skewed. Peptide coverage for kuste3789–96 was 14–58% p.p.d. compared with less than 1% for the other three ATP synthases, and mRNA coverage differed by a factor of six. The gene product encoding the catalytic β -subunit of the highest expressed ATP synthase (kuste3787–96) was recently shown to be associated with the membranes of the intracellular cell compartment, the anammoxosome, suggesting it to be the site where the proton-motive machinery resides²³.

In the present study we experimentally identified NO and N_2H_4 as the intermediates of anaerobic ammonium oxidation. The highly expressed protein encoded by the gene cluster kuste2859–61 was purified and N–N bond formation from NO and ammonium was demonstrated. Hydrazine synthase and the NO reductase of denitrifiers are the two enzymes capable of bonding two N atoms together. In contrast to NO reductase, hydrazine synthase combines two different nitrogenous molecules. It is intriguing that all the N_2 in our atmosphere is formed by the oxidizing power of NO, in line with the hypothesis that NO may have been the first deep redox sink on Earth²⁴.

METHODS SUMMARY

Activity Measurements. Physiological experiments were performed at 33 °C, pH 7.5 with *K. stuttgartiensis* cells^{9,10}. To determine the role of NO and hydroxylamine in the anammox metabolism, cells were incubated with (1) NaNO_2 , NH_4Cl (2 mM each) and spiked with acetylene (15 μM); (2) NaNO_2 , NH_4Cl (2 mM each) and DAF-2DA (10 μM) or PTIO (100 μM); (3) NO (0.1 mM) and 2 mM NH_4Cl . Hydroxylamine, NH_4^+ , NO_2^- and N_2H_4 were determined as previously described^{25,26}. NO was measured online as previously described²⁷. To determine N_2H_4 turnover, cells were incubated with $\text{Na}^{15}\text{NO}_2$, NH_4Cl (or vice versa) and N_2H_4 (2 mM each). Isotopic composition of hydrazine was determined with matrix-assisted laser desorption/ionization–time of flight mass spectrometry (MALDI–TOF MS) after *para*-dimethylaminobenzaldehyde derivatization, developed after Watt and Chrisp²⁵. All labelled compounds were 99% pure (Cambridge Isotope Laboratories).

Proteins were purified from cell-free extracts with anion exchange and hydroxyapatite liquid chromatography. Activity measurements were performed at 37 °C, pH 7 in an anaerobic chamber. Kuste2859–61 (1.6 mg) and kusc1061 (4.7 μg) were incubated with NO (0.9 mM) and $^{15}\text{NH}_4\text{Cl}$ (1 mM), and $^{29}\text{N}_2$ production was monitored by gas chromatography (Agilent 6890 with a PorapakQ column, 80 °C) combined with a mass spectrometer (Agilent 5975c quadrupole inert MS). For rate calculations, kusc0694 (1.3 μg) or kusc1061 (4.7 μg) were

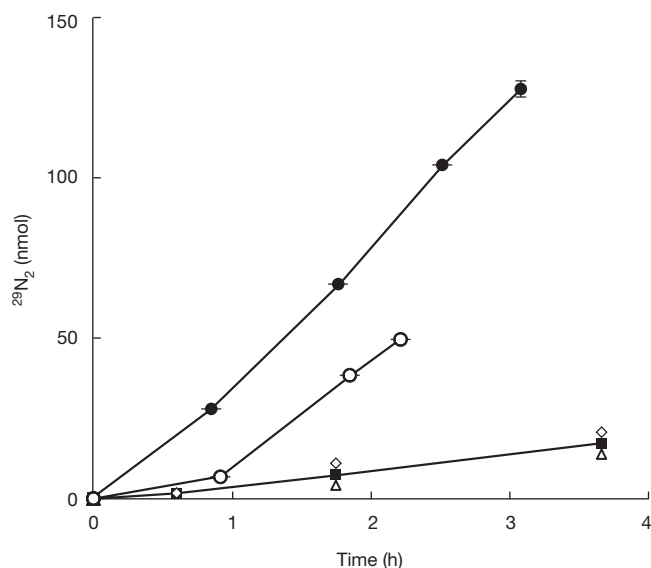


Figure 4 $^{29}\text{N}_2$ production by hydrazine synthase complex and kusc1061 from $^{15}\text{NH}_4^+$ and NO. $^{29}\text{N}_2$ was produced with the highest rate when hydrazine synthase complex (1.6 mg) and kusc1061 (4.7 μg) was incubated with $^{15}\text{NH}_4^+$ (1 mM), ^{14}NO (0.9 mM) and cytochrome *c* (50 μM) (filled circles). In the control experiments, hydrazine synthase complex and cytochrome *c* (open circles), kusc1061 and cytochrome *c* (open diamonds), cytochrome *c* (filled squares) and only buffer (open triangles) were incubated under the same experimental conditions; $n = 3$ (error bars, s.d.).

incubated with N_2H_4 or hydroxylamine and cytochrome *c* (50 μM each), and $^{30}\text{N}_2$ or ^{31}NO production was measured.

Molecular methods. Total RNA was extracted, reverse transcribed, sequenced with Illumina and mapped to the genome sequence of *K. stuttgartiensis*⁶. From the aligned reads, per-position coverage was calculated for each contig and used to calculate the coverage for each orf, intergenic region and predicted RNA element.

Cell free extracts were separated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) or two-dimensional gel electrophoresis, digested with trypsin and analysed with liquid chromatography–mass spectrometry (LC–MS/MS)^{28,29}. Mass spectrometry data was searched against a database of predicted *K. stuttgartiensis* peptide sequences.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 May; accepted 12 August 2011.

Published online 2 October 2011.

- Arrigo, K. R. Marine microorganisms and global nutrient cycles. *Nature* **437**, 349–355 (2005).
- Brandes, J. A., Devol, A. H. & Deutsch, C. New developments in the marine nitrogen cycle. *Chem. Rev.* **107**, 577–589 (2007).
- Payne, W. J. Reduction of nitrogenous oxides by microorganisms. *Bacteriol. Rev.* **37**, 409–452 (1973).
- Strous, M. *et al.* Missing lithotroph identified as new planctomycete. *Nature* **400**, 446–449 (1999).
- Zumft, W. G. Cell biology and molecular basis of denitrification. *Microbiol. Mol. Biol. Rev.* **61**, 533–616 (1997).
- Strous, M. *et al.* Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**, 790–794 (2006).
- Van de Graaf, A. A., deBruijn, P., Robertson, L. A., Jetten, M. S. M. & Kuenen, J. G. Metabolic pathway of anaerobic ammonium oxidation on the basis of N-15 studies in a fluidized bed reactor. *Microbiology* **143**, 2415–2421 (1997).
- Kartal, B. *et al.* Candidatus ‘Brocadia fulgida’: an autofluorescent anaerobic ammonium oxidizing bacterium. *FEMS Microbiol. Ecol.* **63**, 46–55 (2008).
- Kartal, B., Geerts, W. & Jetten, M. S. M. Cultivation, detection and ecophysiology of anaerobic ammonium-oxidizing bacteria. *Methods Enzymol.* **486**, 89–108 (2011).
- Van der Star, W. R. L. *et al.* The membrane bioreactor: a novel tool to grow anammox bacteria as free cells. *Biotechnol. Bioeng.* **101**, 286–294 (2008).
- Keller, M. & Hettich, R. Environmental proteomics: a paradigm shift in characterizing microbial activities at the molecular level. *Microbiol. Mol. Biol. Rev.* **73**, 62–70 (2009).
- Akaike, T. & Maeda, H. Quantitation of nitric oxide using 2-phenyl-4,4,5,5-tetramethylimidazole-1-oxyl 3-oxide (PTIO). *Methods Enzymol.* **268**, 211–221 (1996).
- Guo, F. Q., Okamoto, M. & Crawford, N. M. Identification of a plant nitric oxide synthase gene involved in hormonal signaling. *Science* **302**, 100–103 (2003).
- Nagano, T. Practical methods for detection of nitric oxide. *Luminescence* **14**, 283–290 (1999).
- Gilch, S., Vogel, M., Lorenz, M. W., Meyer, O. & Schmidt, I. Interaction of the mechanism-based inactivator acetylene with ammonia monooxygenase of *Nitrosomonas europaea*. *Microbiology* **155**, 279–284 (2009).
- Hyman, M. R. & Wood, P. M. Suicidal inactivation and labeling of ammonia monooxygenase by acetylene. *Biochem. J.* **227**, 719–725 (1985).
- McTavish, H., Fuchs, J. A. & Hooper, A. B. Sequence of the gene coding for ammonia monooxygenase in *Nitrosomonas europaea*. *J. Bacteriol.* **175**, 2436–2444 (1993).
- Hooper, A. B. & Nason, A. Characterization of hydroxylamine-cytochrome *c* reductase from chemoautotrophs *Nitrosomonas europaea* and *Nitrosocystis oceanus*. *J. Biol. Chem.* **240**, 4044–4057 (1965).
- Hooper, A. B., Vannelli, T., Bergmann, D. J. & Arciero, D. M. Enzymology of the oxidation of ammonia to nitrite by bacteria. *Anton. Leeuw. Int. J. G. Microbiology* **71**, 59–67 (1997).
- Klotz, M. G. *et al.* Evolution of an octahaem cytochrome *c* protein family that is key to aerobic and anaerobic ammonia oxidation by bacteria. *Environ. Microbiol.* **10**, 3150–3163 (2008).
- Shimamura, M. *et al.* Isolation of a multiheme protein with features of a hydrazine-oxidizing enzyme from an anaerobic ammonium-oxidizing enrichment culture. *Appl. Environ. Microbiol.* **73**, 1065–1072 (2007).
- Shimamura, M. *et al.* Another multiheme protein, hydroxylamine oxidoreductase, abundantly produced in an anammox bacterium besides the hydrazine-oxidizing enzyme. *J. Biosci. Bioeng.* **105**, 243–248 (2008).
- Van Niftrik, L. *et al.* Intracellular localization of membrane-bound ATPases in the compartmentalized anammox bacterium ‘Candidatus Kuenenia stuttgartiensis’. *Mol. Microbiol.* **77**, 701–715 (2010).
- Ducluzeau, A. L. *et al.* Was nitric oxide the first deep electron sink? *Trends Biochem. Sci.* **34**, 9–15 (2009).
- Watt, G. W. & Crisp, J. D. A spectrophotometric method for the determination of hydrazine. *Anal. Chem.* **24**, 2006–2008 (1952).
- Strous, M., Heijnen, J. J., Kuenen, J. G. & Jetten, M. S. M. The sequencing batch reactor as a powerful tool for the study of slowly growing anaerobic ammonium-oxidizing microorganisms. *Appl. Microbiol. Biotechnol.* **50**, 589–596 (1998).
- Kartal, B. *et al.* Effect of nitric oxide on anammox bacteria. *Appl. Environ. Microbiol.* **76**, 6304–6306 (2010).
- Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
- Wilm, M. *et al.* Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466–469 (1996).
- Audrieth, L. F. & Ackerson Ogg, B. *The Chemistry of Hydrazine* (Wiley, 1951).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements B.K. was supported by a grant (05987) from the Dutch Foundation for Applied Research. W.J.M. was supported by a grant (142161201) from the Darwin Center for Biogeosciences. N.M.d.A. and I.C. were supported by a grant (81802015) from the Netherlands Organization for Scientific Research. M.S. was supported by a VIDI grant from the Netherlands Organization for Scientific Research and a European Research Council grant MASEM (242635). The anammox research of M.S.M.J. is supported by an advanced grant (232987) from the ERC. The authors acknowledge R. Klefoth for the initial tests for protein purification procedures.

Author Contributions Physiological experiments were conceived, designed and performed by B.K., *Kuenenia stuttgartiensis* was grown by B.K. and W.G., two-dimensional gel electrophoresis was performed by N.M.A. and I.C., one-dimensional gel electrophoresis was performed by W.J.M. and B.K., MALDI-TOF analysis was performed by B.K., W.J.M. and H.J.M.O.d.C., nanoLC-MS/MS by J.G., RNA extraction and reverse transcription by H.R.H., Illumina sequencing by E.M.J.-M., K.-J.F. and H.S., and protein purification and activity tests were designed by W.J.M., B.K. and J.T.K. and performed by W.J.M. Proteomic and transcriptomic data processing was performed by J.G., M.S., K.-J.F., B.K., M.S.M.J. and H.J.M.O.d.C. The manuscript was written by B.K. with input from J.T.K., M.S. and M.S.M.J.

Author Information The metatranscriptome and peptidome sequences are deposited in Gene Expression Omnibus under accession numbers GSE15408 and PSE111, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.K. (kartal@science.ru.nl).

METHODS

Source of the biomass. *K. stuttgartiensis* cells were collected from a 10-l laboratory scale anammox membrane bioreactor^{9,10} and were concentrated by centrifugation. The cells were re-suspended to a protein concentration higher than 1 mg ml⁻¹. Part of the cell suspension was diluted 100 times, chemically fixed, and hybridizations with fluorescently labelled oligonucleotide probes were performed as described previously^{31,32}.

Sample preparation. The cell suspensions were transferred to 8-ml serum bottles. The vials were made anoxic by alternately applying under-pressure and He or Ar seven times and were transferred to an anaerobic chamber with a 95%/5% Ar/H₂ atmosphere. O₂ in the Ar in the anaerobic chamber was removed by passing Ar over a Pd catalyst (0.2 p.p.m. residual O₂). In the anaerobic chamber, cell suspensions were diluted five times with anaerobic mineral medium³³ (pH 7.5) to a final volume of 40 or 8 ml and transferred to glass vials unless stated otherwise. All preparations (for example, addition of substrates and/or inhibitors) for different incubations were handled in the anaerobic chamber. All experiments were performed at least in duplicate. All non-labelled salts were purchased as molecular grade (more than 99.95% pure, Merck) unless stated otherwise. All labelled compounds were 99% pure and purchased as sodium or chloride salts (Cambridge Isotope Laboratories). All gaseous compounds were of the highest purity available.

Analytical methods. NO₂⁻ and NH₄⁺ were determined as described previously²⁷. N₂H₄ was determined colourimetrically at 420 nm after reaction of 100-μl sample with 900 μl 2% (w/v) *para*-dimethylaminobenzaldehyde (PDB), 3.7% (v/v) HCl in ethanol³⁰. NH₂OH (detection limit 5 μM) was determined as previously described²⁷.

Effect of PTIO. To determine the effect of PTIO, an NO scavenger¹², on anammox bacteria, three incubations were performed in parallel. NO₂⁻ and NH₄⁺ (2 mM) were added to all incubations. To the first incubation, PTIO (100 μM) was added at 0 min, to the second it was added at 40 min, and no PTIO was added to the third incubation. Liquid samples were taken every 15 min and analysed for NH₄⁺, NO₂⁻ and NH₂OH as previously described²⁷.

Bioimaging of nitric oxide. To detect NO turnover, *K. stuttgartiensis* cell suspensions were incubated with 2 mM NO₂⁻ and NH₄⁺ in amber vials. In parallel, nitrite-depleted cell suspensions were incubated in the presence of 2 mM NH₄⁺. After a 5-min pre-incubation, diamino fluorescein-2-diacetate (DAF2-DA, Calbiochem) was added to a final concentration of 10 μM. The vials were incubated in the dark for 30 min at 33 °C and were shaken continuously at 300 r.p.m. As a negative control, cells were incubated with PTIO and DAF2-DA. Cells were then harvested by centrifugation, washed three times in mineral medium³³ to remove the excess chromophore and were re-suspended in mineral medium. A liquid sample (5 μl) of the suspension was pipetted on a microscope slide and dried in the dark. The preparations were examined with a Zeiss Axioplan2 epifluorescence microscope.

Batch experiments. To determine the activity of *K. stuttgartiensis* with NO and NH₄⁺, cell suspensions were incubated with NO (0.1 mM) and 2 mM NH₄⁺ in 100-ml glass vials with 10% NO (in He) in the headspace. Gas samples were analysed in a chemoluminescence NO_x analyser (CLD 700EL, EcoPhysics, detection limit 0.1 p.p.m. NO). Liquid samples were taken every 30 min and analysed for NH₄⁺ and N₂H₄ as previously described^{27,30}.

To determine the effect of acetylene on anammox bacteria, *K. stuttgartiensis* suspensions were transferred to 40-ml glass vials. NO₂⁻ and NH₄⁺ were added to the incubations to a final concentration of 2 mM. The vials were incubated at 33 °C and were mixed with a magnetic stirrer at 500 r.p.m. and continuously flushed with Ar/CO₂ (95%/5%) with a flow of 10 ml min⁻¹. The effluent gas from the vials was connected to a chemoluminescence NO_x analyser (CLD 700EL, EcoPhysics, detection limit 0.1 p.p.m. NO) for online NO measurement. At 15 min, 100 μl acetylene (15 μM) was added to the vials. As negative controls, 100 μl air and 100 μl nitrogen were added to separate incubations. Liquid samples were taken from the incubations every 10–15 s and chilled to 0 °C immediately. The supernatant of each sample was transferred to an Eppendorf cup and kept at 4 °C until they were analysed for NH₄⁺, N₂H₄ and NH₂OH.

Detection of hydrazine turnover in anammox cells. To detect N₂H₄ turnover, ¹⁵NO₂⁻, NH₄⁺ (or vice versa) and N₂H₄ (2 mM each) were added to the *K. stuttgartiensis* cell suspensions. The vials were incubated in the dark for 15 min at 30 °C, 300 r.p.m. Liquid samples were taken every 5 min, and isotopic composition of N₂H₄ was determined by MALDI-TOF MS after reaction with PDB. For MALDI-TOF analysis, 10 μl of PDB-reacted samples were mixed with an equal volume of sample buffer containing 20 mg ml⁻¹ α-cyano-4-hydroxycinnamic acid in 0.05% (v/v) trifluoroacetic acid (TFA), 50% (v/v) acetonitrile. The mixtures (0.3 μl) were spotted on a S26/100 M-probe (Bruker 15165), which was inserted into a multiprobe adaptor. MALDI-TOF MS measurements were performed in the mass range of 100–800 Da on a Bruker III mass spectrometer, using the reflectron mode.

Cytochrome *bc*₁ complex. To determine the role of cytochrome *bc*₁ complex in anammox catabolism, *K. stuttgartiensis* suspensions were incubated with pentachlorophenol, a specific inhibitor of the *bc*₁ complex. NO₂⁻ and NH₄⁺ were added to the incubations to a final concentration of 2 mM. Pentachlorophenol was added to a final concentration of 10 μM. NO₂⁻ and NH₄⁺ were determined as described previously²⁷.

Preparation of cell free extract. *K. stuttgartiensis* cells (2 l, OD₆₀₀ 1.2) were harvested from the membrane bioreactor. After centrifugation (4,000g, 4 °C), the pellet was re-suspended in one volume 20 mM potassium phosphate buffer, pH 8. Cell suspensions were passed three times through a French pressure cell operated at 138 MPa. The lysate was incubated with 1% (w/v) sodium deoxycholate at 4 °C for 1 h to solubilize membrane associated proteins. After centrifugation for 15 min at 1,700g at 4 °C, the cell-free fraction was obtained as clarified supernatant.

Protein electrophoresis and MALDI-TOF analysis. Samples were denatured by incubation with 60 mM Tris-HCl buffer (pH 8) containing 5% β-mercaptoethanol, 2% SDS (sodium dodecyl sulphate) and 25% glycerol for 5 min at 100 °C. SDS-PAGE was performed in 10% or 6% slab gels in 375 mM Tris-HCl glycine buffer, pH 8.8 according to Laemmli³⁴. Native PAGE (6%) was performed according to the same procedure with the following modifications: the protein preparations were not boiled before electrophoresis, SDS and β-mercaptoethanol were omitted from the gels, and Tris-HCl glycine (375 mM, pH 8.3) was used as the running buffer. Gels were stained with colloidal Coomassie blue as described elsewhere³⁵. To identify the protein bands resolved in SDS-PAGE, gel spots (~3 mm²) were picked, digested with trypsin and analysed with MALDI-TOF mass spectrometry as described elsewhere³⁶.

Purification of kuste2859–2860–2861, kusc0694 and kusc1061. Cell-free extract was centrifuged at 140,000g, 10 °C (Discovery 10, Sorvall, equipped with a T-1270 rotor) to remove the membranes. The supernatant was loaded on a 30 ml Q Sepharose XL (GE Healthcare) column equilibrated with 20 mM Tris-HCl, pH 8. Kuste2859–2860–2861 and kusc1061 were eluted isocratically with 200 mM NaCl in 20 mM Tris-HCl, pH 8 (2 ml min⁻¹). Kusc0694 was eluted isocratically with 400 mM NaCl in 20 mM Tris-HCl, pH 8 (2 ml min⁻¹). Eluted fractions were subsequently loaded onto a 10 ml Hydroxapate (Bio-Rad) column equilibrated with 20 mM potassium phosphate buffer, pH 7 and eluted with a gradient of the same buffer (20–500 mM, 2 ml min⁻¹). Kusc1061 and kuste2859–2860–2861 were collected in fractions eluted at 100 mM and 200 mM phosphate, respectively. The pooled fractions were desalted and concentrated using Vivaspinn tubes (100 kDa cut-off, Sartorius Stedim Biotech) to concentrations of at least 0.86 mg ml⁻¹ (kuste2859–2860–2861) and 2.3 mg ml⁻¹ (kusc1061) in 20 mM phosphate buffer, pH 7.

Detection of hydrazine and hydroxylamine oxidation by kusc1061 and kusc0694. To 2 ml (final volume) of phosphate buffer (20 mM, pH 7), 4.7 μg of Kusc1061 or 1.3 μg of Kusc0694 and cytochrome *c* (50 μM final concentration, bovine heart, Sigma-Aldrich) were added to a 3-ml extainer (Labco). To start the reaction to determine the electron stoichiometry, 10 μM, and for routine rate assays 50 μM, ¹⁵N-labelled ³⁰N₂H₄ was added from an anoxic stock. To determine the capacity for NH₂OH oxidation, proteins were incubated in separate vials with 50 μM NH₂OH and cytochrome *c* (each). Extainers were incubated at 37 °C in the anaerobic chamber. ³⁰N₂ and ¹⁵NO production was monitored by gas chromatography (Agilent 6890 equipped with a Porapak Q column at 80 °C) combined with a mass spectrometer (Agilent 5975c quadrupole inert MS).

Combined assay of kuste2859–2860–2861 and kusc1061. Cytochrome *c* (50 μM final concentration, bovine heart, Sigma-Aldrich), Kusc1061 (4.7 μg), 1 mM ¹⁵NH₄⁺ and 5 μM N₂H₄ were added to 1.6 mg of kuste2859–2860–2861 in 1 ml phosphate buffer (20 mM, pH 7) in a 3-ml extainer (Labco). The reaction was started by adding phosphate buffer (20 mM, pH 7) with NO (0.9 mM) to a final volume of 2 ml. Before incubation at 37 °C in the anaerobic chamber, 1 ml of 50% NO (in He) was added to the headspace. Control experiments were performed with ammonium (1 mM) with NH₂OH (1 mM) and HNO supplied as Angeli's salt (41 mM) in separate incubations. ²⁹N₂ production was monitored by gas chromatography (Agilent 6890 equipped with a Porapak Q column at 80 °C) combined with a mass spectrometer (Agilent 5975c quadrupole inert MS).

LC-MS/MS analysis and data processing. After PAGE, gels were stained with colloidal Coomassie blue as described elsewhere³⁵. The gel lane was cut into four slices and each slice was destained with three cycles of washing with 50 mM ammoniumbicarbonate and 50% acetonitrile. Protein reduction, alkylation and digestion with trypsin were performed as previously described³⁰. After digestion, samples were de-salted and purified according to Rappsilber *et al.*²⁹. Sample analysis by LC-MS/MS was performed using an Agilent nanoflow 1100 liquid chromatograph coupled online through a nano-electrospray ion source (Thermo Fisher Scientific) to a 7T linear ion trap Fourier transform ion cyclotron resonance mass spectrometer (LTQ FT, Thermo Fisher Scientific). The chromatographic column consisted of a 15-cm fused silica emitter (New Objective, PicoTip

Emitter, Tip: $8 \pm 1 \mu\text{m}$, internal diameter $100 \mu\text{m}$) packed with $3\text{-}\mu\text{m}$ C18 beads (Reprosil-Pur C18 AQ, Dr Maisch GmbH)³⁷. After loading the peptides onto the column in buffer A (0.5% HAc), bound peptides were gradually eluted using a 67-min gradient of buffer B (80% ACN, 0.5% HAc). First, the concentration of acetonitrile was increased from 2.4 to 8% in 5 min, followed by an increase from 8 to 24% acetonitrile in 55 min, and finally an increase from 24 to 40% acetonitrile in 7 min. The mass spectrometer was operated in positive ion mode and was programmed to analyse the top four most abundant ions from each precursor scan using dynamic exclusion. Survey mass spectra (350–2000 m/z) were recorded in the ion cyclotron resonance cell at a resolution of $R = 5\text{E}5$. Data-dependent collision-induced fragmentation of the precursor ions was performed in the linear ion trap (normalized collision energy 27%, activation $q = 0.250$, activation time 30 ms).

Mass spectrometric data files were searched against the *K. stuttgartiensis* database (known contaminants like human keratins and trypsin were added to the database) using the database search program Mascot (Matrix Science, version 2.2). To obtain factors for the recalibration of precursor masses, initial searches were performed with a precursor ion tolerance of 50 p.p.m. Fragment ions were searched with 0.8-Da tolerance and searches allowed for one missed cleavage, carbamidomethylation (C) as fixed modification, and deamidation (NQ) and oxidation (M) as variable modifications. The results from these searches were used to calculate the m/z -dependent deviation, which was used to recalibrate all precursor m/z values. After recalibration of the precursor masses, definitive Mascot searches were performed using the same settings as stated above, but with a precursor ion tolerance of 20 p.p.m. Additionally, reverse database searches were performed with the same settings. Protein identifications were validated and clustered using the PROVALT algorithm to achieve a false-discovery rate of less than 1% (ref. 38).

Two-dimensional gel electrophoresis. Before protein separation by isoelectric focusing, 1 mg of the protein suspension was incubated with 1% (v/v) Immobilized pH-gradient (IPG) buffer of the appropriate range, 5 mM tributyl phosphine and 0.01% (w/v) bromophenol blue for 15 min at room temperature and centrifuged at 10,000g for 15 min at 10°C . Isoelectric focusing was performed with the IPGphor system using commercial 24-cm-long IPG strips with linear immobilized pH gradients of various ranges. The conditions for rehydration of the IPG strips, sample entry and isoelectric focusing were as follows: the temperature was set constant at 18°C and 50- μA per strip were applied.

Focused IPG strips were equilibrated before SDS-PAGE two times for 15 min in 375 mM Tris-HCl pH 8.5, 2% (w/v) SDS, 20% (w/v) glycerol, 6 M urea, 10 mM DTT, 50 mM acrylamide and 0.1% (w/v) bromophenol blue. Gels were run for 45 min with constant cooling to 18°C at 20 V, 40 W and subsequently at 40 V, 40 W until the bromophenol blue marker reached the end of the gel. Gels were fixed in 30% (v/v) ethanol and 10% (v/v) glacial acetic acid and were stained with colloidal Coomassie blue³⁹ or silver stain⁴⁰. Picked gel spots were digested and analysed with MALDI-TOF MS as described elsewhere³⁶.

Blue native PAGE. Blue native PAGE of the protein complexes was performed as described elsewhere⁴¹. For protein identification in two-dimensional gels, $16\text{ cm} \times 20\text{ cm}$ gels were self-casted according to Calvaruso *et al.* with the following

exception: 4–10% linear polyacrylamide gradient was used⁴¹. Sample additive (1.5 μl) (0.75 M 6-aminocaproic acid, 5% Serva Blue G) was added to 40 μg protein sample before loading the gel.

Electrophoresis was performed at 50 V until the migration front entered the resolving gel and then at 100 V until the migration front reached the end of the gel. Cathode and anode buffer for blue native PAGE were 50 mM Bis-Tris, pH 7.0, and 50 mM Tricine, 15 mM Bis-Tris, pH 7.0, respectively. Preparation of the first-dimension gel strip and assembly and casting of the second-dimension gel were performed as described elsewhere⁴¹ with the exception that the second-dimension cassette had the same thickness as the first dimension. No Coomassie blue was added to the cathode buffer.

Transcriptomics. RNA was extracted using the Ribopure Bacteria Kit (Ambion) according to the manufacturer's instructions. First-strand cDNA was synthesized with random primers using the RevertAid H Minus First Strand cDNA Synthesis Kit, and the second strand was synthesized using DNA polymerase and manufacturer's instructions (Fermentas).

The quality scores of the obtained Solexa reads (3.5 million) were converted to PHRED format and mapped with Maq (<http://maq.sourceforge.net>) to the five contigs that constitute the *K. stuttgartiensis* genome (accession numbers CT030148, CT573071–4). From the aligned reads, the per-position coverage was calculated for each contig and used to calculate the coverage for each orf, intergenic region and predicted RNA element.

31. Schmid, M. *et al.* Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. *Syst. Appl. Microbiol.* **23**, 93–106 (2000).
32. Schmid, M. C. *et al.* Biomarkers for in situ detection of anaerobic ammonium-oxidizing (anammox) bacteria. *Appl. Environ. Microbiol.* **71**, 1677–1684 (2005).
33. Van de Graaf, A. A., de Bruijn, P., Robertson, L. A., Jetten, M. S. M. & Kuenen, J. G. Autotrophic growth of anaerobic ammonium-oxidizing micro-organisms in a fluidized bed reactor. *Microbiology* **142**, 2187–2196 (1996).
34. Laemmli, U. K. Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature* **227**, 680–685 (1970).
35. Candiano, G. *et al.* Blue silver: A very sensitive colloidal Coomassie G-250 staining for proteome analysis. *Electrophoresis* **25**, 1327–1333 (2004).
36. Farhoud, M. H. *et al.* Protein complexes in the archaeon *Methanothermobacter thermotrophicus* analyzed by blue native/SDS-PAGE and mass spectrometry. *Mol. Cell. Proteomics* **4**, 1653–1663 (2005).
37. Ishihama, Y., Rappsilber, J., Andersen, J. S. & Mann, M. Microcolumns with self-assembled particle frits for proteomics. *J. Chromatogr. A* **979**, 233–239 (2002).
38. Weatherly, D. B. *et al.* A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Mol. Cell. Proteomics* **4**, 762–772 (2005).
39. Neuhoff, V., Arold, N., Taube, D. & Ehrhardt, W. Improved staining of proteins in polyacrylamide gels including isoelectric-focusing gels with clear background at nanogram sensitivity using coomassie brilliant blue G-250 and R-250. *Electrophoresis* **9**, 255–262 (1988).
40. Mortz, E., Krogh, T. N., Vorum, H. & Gorg, A. Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics* **1**, 1359–1363 (2001).
41. Calvaruso, M. A., Smeitink, J. & Nijtmans, L. Electrophoresis techniques to investigate defects in oxidative phosphorylation. *Methods* **46**, 281–287 (2008).

Metabolic priming by a secreted fungal effector

Armin Djamei^{1*}, Kerstin Schipper^{1,2*†}, Franziska Rabe¹, Anupama Ghosh¹, Volker Vincon¹, Jörg Kahnt¹, Sonia Osorio³, Takayuki Tohge³, Alisdair R. Fernie³, Ivo Feussner⁴, Kirstin Feussner⁵, Peter Meinicke⁵, York-Dieter Stierhof⁶, Heinz Schwarz⁷, Boris Macek^{8†}, Matthias Mann⁸ & Regine Kahmann¹

Maize smut caused by the fungus *Ustilago maydis* is a widespread disease characterized by the development of large plant tumours. *U. maydis* is a biotrophic pathogen that requires living plant tissue for its development and establishes an intimate interaction zone between fungal hyphae and the plant plasma membrane. *U. maydis* actively suppresses plant defence responses by secreted protein effectors^{1,2}. Its effector repertoire comprises at least 386 genes mostly encoding proteins of unknown function^{1,3,4} and expressed exclusively during the biotrophic stage³. The *U. maydis* secretome also contains about 150 proteins with probable roles in fungal nutrition, fungal cell wall modification and host penetration as well as proteins unlikely to act in the fungal-host interface⁴ like a chorismate mutase. Chorismate mutases are key enzymes of the shikimate pathway and catalyse the conversion of chorismate to prephenate, the precursor for tyrosine and phenylalanine synthesis. Root-knot nematodes inject a secreted chorismate mutase into plant cells likely to affect development^{5,6}. Here we show that the chorismate mutase Cmu1 secreted by *U. maydis* is a virulence factor. The enzyme is taken up by plant cells, can spread to neighbouring cells and changes the metabolic status of these cells through metabolic priming. Secreted chorismate mutases are found in many plant-associated microbes and might serve as general tools for host manipulation.

The *U. maydis* genome (<http://mips.helmholtz-muenchen.de/genet/proj/ustilago>) contains genes for both a cytosolic chorismate mutase, designated *aro7* (*um04220*), and a putatively secreted chorismate mutase, *cmu1* (*um05731*). Cmu1 belongs to the AroQ class of eukaryotic chorismate mutases (Interpro: IPR008238) that have an all- α -helical secondary structure (Supplementary Fig. 1)^{7,8}. To verify that Cmu1 is a dedicated chorismate mutase, we demonstrated that it complemented a *Saccharomyces cerevisiae* *aro7* mutant (Fig. 1a) and that heterologously expressed protein had chorismate mutase activity which was not feedback inhibited by aromatic amino acids (Supplementary Fig. 2). Allosteric regulation is a characteristic feature of plastidic chorismate mutases as well as of cytoplasmic fungal chorismate mutases^{9,10}, whereas cytosolic plant chorismate mutases lack this feature¹¹. Attempts to generate a *cmu1* mutant that displayed allosteric regulation based on features of *S. cerevisiae* Aro7p were unsuccessful (Supplementary Fig. 3). Western blot analysis detected Cmu1-haemagglutinin (HA) in *U. maydis* culture supernatants when the respective fusion gene was expressed under a constitutive promoter in hyphal cells (Supplementary Fig. 4). The secretion of Cmu1 during plant colonization was independently demonstrated by proteome analysis of apoplastic fluids isolated after infection of maize with a mixture of compatible *U. maydis* strains. Compared with known secreted effector proteins like Mig2 (ref. 12), higher numbers of Cmu1 peptides were identified at all time points analysed (Supplementary Information, Table 1).

Like many other *U. maydis* effectors with a virulence function^{1,3,13}, *cmu1* is specifically upregulated during biotrophic development (Supplementary Fig. 5) and is one of the most highly expressed fungal genes during plant colonization¹⁴. To determine a possible contribution to virulence *cmu1* was deleted in the solopathogenic strains SG200 (ref. 3) and CL13. CL13 is the progenitor strain of SG200 that shows attenuated virulence¹⁵ (see Supplementary Fig. 6a for disease

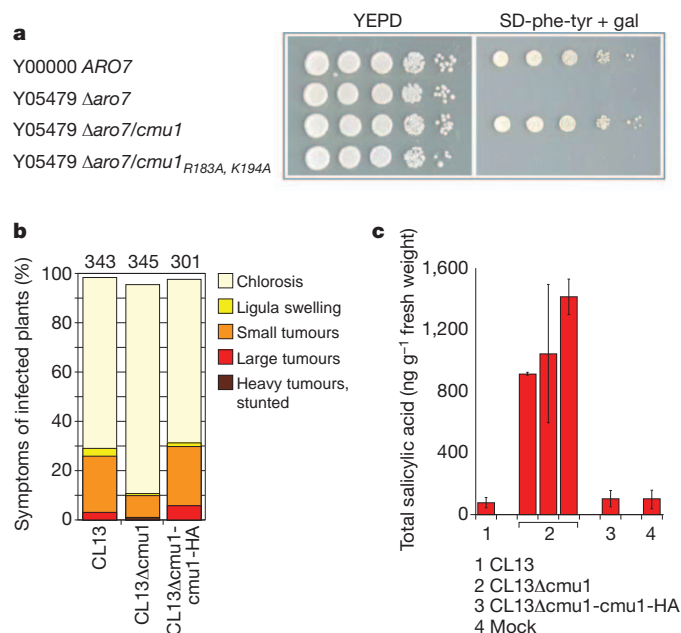


Figure 1 | Cmu1 has chorismate mutase activity, affects virulence and salicylic acid levels. **a**, *U. maydis* *cmu1* complements the *aro7* deletion in *S. cerevisiae*. Growth of the *S. cerevisiae* *aro7* deletion mutant Y05479 on medium lacking tyrosine and phenylalanine (SD-phe-tyr) is restored by introduction of *U. maydis* *cmu1*; whereas *cmu1*_{R183A, K194A} does not complement. Expression of *cmu1* genes was driven by the *GAL1* promoter. *S. cerevisiae* Y00000 (native ARO7 gene) was used as positive control. YEPD, rich medium; gal, galactose. **b**, Deletion of *cmu1* negatively affects virulence of *U. maydis* strain CL13. Disease symptoms (as depicted in Supplementary Fig. 6a) on maize plants were scored 12 days after infection with the indicated strains. Mean values of seven independent infections are shown with the total number of infected plants indicated above each column. Compared with CL13 and CL13 Δ cmu1-cmu1-HA, CL13 Δ cmu1 showed significantly reduced tumour formation (*t*-test, *P* = 0.037). **c**, Total amounts of salicylic acid were determined in plant leaves infected with the indicated *U. maydis* strains listed below 8 days after infection. For the infections with CL13 Δ cmu1, three independent strains were used. Mean values of three independent experiments are shown. Error bars, s.d.

¹Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Straße 10, D-35043 Marburg, Germany. ²Zentrum Synthetische Mikrobiologie, Philipps-Universität Marburg, D-35032 Marburg, Germany. ³Max Planck Institute of Molecular Plant Physiology, Wissenschaftspark Golm, Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany. ⁴Georg-August-Universität, Albrecht-von-Haller Institute, Justus-von-Liebig Weg 11, D-37077 Göttingen, Germany. ⁵Georg-August-Universität, Institute for Microbiology and Genetics, Grisebachstraße 8, D-37077 Göttingen, Germany. ⁶Center for Plant Molecular Biology, University of Tübingen, Auf der Morgenstelle 5, D-72076 Tübingen, Germany. ⁷Max Planck Institute for Developmental Biology, Spemannstraße 35, D-72076 Tübingen, Germany. ⁸Max Planck Institute for Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany. [†]Present addresses: Heinrich-Heine-Universität, Abteilung Mikrobiologie, Gebäude 26.12, Ebene 01, Universitätsstrasse 1, D-40225 Düsseldorf, Germany (K.S.); Proteome Center Tuebingen, Auf der Morgenstelle 15, D-72076 Tübingen, Germany (B.M.).

*These authors contributed equally to this work.

symptoms of CL13) and hence facilitates the detection of modest differences in virulence¹⁶. Whereas SG200 Δ cmu1 strains showed little virulence attenuation (Supplementary Fig. 6b), the CL13 Δ cmu1 mutant displayed a reduction of about 50% in tumours, which could be complemented by introducing a single copy of *cmu1*-HA (Fig. 1b). This illustrates that Cmu1 is required for full virulence and demonstrates functionality of the HA-tagged protein.

To localize Cmu1 during biotrophic growth, plants were infected with SG200 Δ cmu1-cmu1-HA, which carries a *cmu1*-HA fusion gene inserted in single copy under control of its native promoter. Plants infected with SG200 or with SG200 P_{cmu1}GFP-HA expressing cytoplasmic green fluorescent protein (GFP) under the *cmu1* promoter served as negative controls. Freeze-substituted and resin-embedded sections of maize tissue harvested 3 days after infection with these strains were incubated with anti-HA antibodies and gold markers. Cmu1-HA could be detected inside the fungal hyphae, in the biotrophic interface as well as inside the plant cytoplasm but rarely in the plant cell wall (Fig. 2A and Supplementary Fig. 7). The distribution of gold particles was quantified (Fig. 2B). Gold labelling of plant tissue infected with the parental strain SG200 was negligible (Supplementary Fig. 8), whereas non-secreted GFP-HA was absent from the biotrophic interphase, showed strong accumulation in the fungal cytosol and weak background labelling in the plant cytosol (Supplementary Fig. 9 and Fig. 2B). Integrity of Cmu1-HA was demonstrated by western blot analysis after immunoprecipitation from infected plant tissue (Supplementary Fig. 10). To demonstrate Cmu1 localization independently, plants were infected with SG200 Δ cmu1-cmu1-mCherry-HA. Cmu1-mCherry-HA was detected in the biotrophic interface, and plasmolysis experiments showed that it freely diffused in the enlarged apoplast (Supplementary Fig. 11). However, fluorescence could not be detected inside plant cells. In addition, Cmu1-mCherry-HA was unable to complement the virulence phenotype of CL13 Δ cmu1 (Supplementary Fig. 12a) despite the fact that the fusion protein was enzymatically active as demonstrated by complementation of the *aro7* yeast mutant (Supplementary Fig. 13). Cmu1₂₂₋₂₉₀-HA lacking the secretion signal was unable to complement the virulence phenotype of CL13 Δ cmu1 (Supplementary Fig. 12b), demonstrating that secretion is prerequisite for function. In sum, these data suggest that Cmu1 needs to enter plants cells to exert its function and that Cmu1-mCherry-HA is unable to do so.

To elucidate the subcellular localization of Cmu1 in plant cells, a Cmu1₂₂₋₂₉₀-mCherry fusion protein lacking the signal peptide and active in complementing the yeast *aro7* mutant was transiently expressed in maize leaves (Supplementary Fig. 13). Cmu1₂₂₋₂₉₀-mCherry localized to the cytoplasm and the nucleus of transformed maize cells (Fig. 2C). Surprisingly, in some cases the Cmu1₂₂₋₂₉₀-mCherry signal was also visible in cells adjacent to the originally transformed cell (Fig. 2C, a). To rule out that the latter is caused by independent transformation events, Cmu1₂₂₋₂₉₀-yellow fluorescent protein (YFP) and PIP₄₂₆₋₅₉₃-mCherry encoding a fusion protein that localizes exclusively to the nucleus, were co-expressed (Supplementary Fig. 14). Cell-to-cell spreading of Cmu1₂₂₋₂₉₀-YFP was observed in some cases whereas PIP₄₂₆₋₅₉₃-mCherry always remained in the nucleus of the originally transformed cell (Supplementary Fig. 14). Occasionally guard cells were transformed, and in such cases spreading of Cmu1₂₂₋₂₉₀-YFP was never observed (Fig. 2C, c and Supplementary Fig. 14). Because guard cells lack plasmodesmata¹⁷, the observed spreading of Cmu1₂₂₋₂₉₀ is likely to occur through plasmodesmata.

By yeast two-hybrid analysis we demonstrated that Cmu1 can dimerize (Supplementary Fig. 15a), a property characteristic for AroQ chorismate mutases⁸. In addition, Cmu1 could interact with the two maize chorismate mutases ZmCm1 (B6TU00) and ZmCm2 (B4FUP5) (Supplementary Fig. 15a). Despite low overall sequence conservation, known residues essential for chorismate mutase activity were conserved in all these enzymes (Supplementary Fig. 15b).

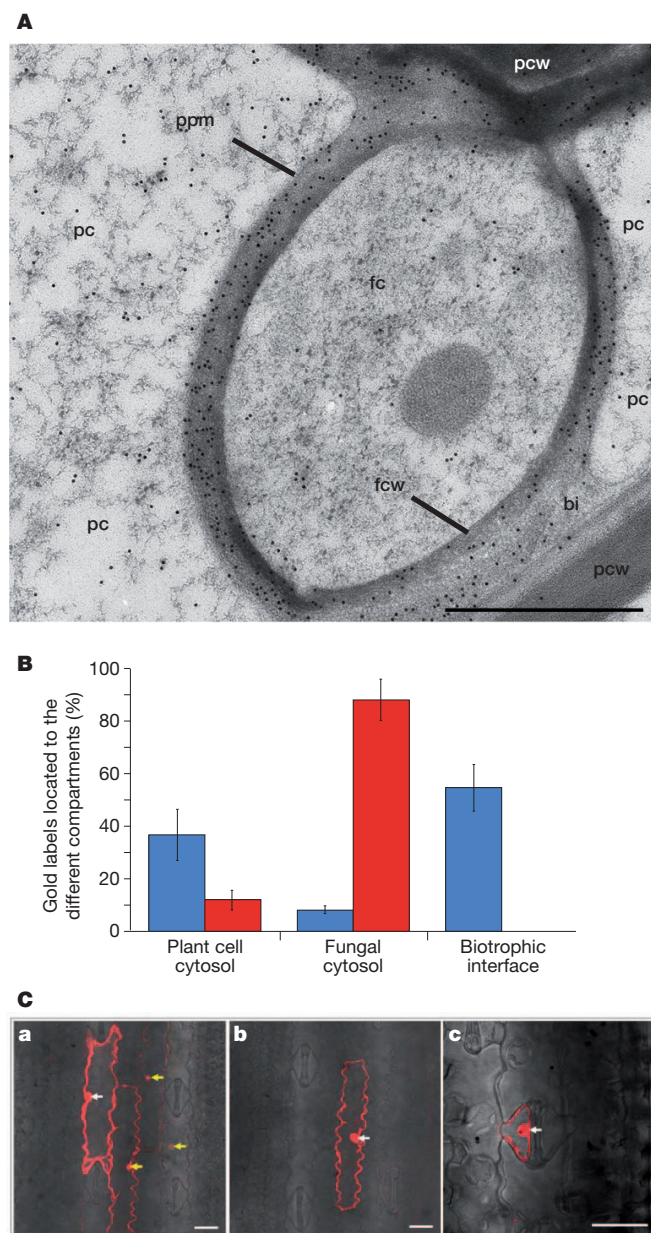


Figure 2 | Cmu1 is translocated to plant cells and spreads to neighbouring tissue. **A**, A maize section infected with *U. maydis* SG200 Δ cmu1-cmu1-HA was collected 3 days after infection, probed with mouse anti-HA antibodies and detected with anti-mouse antibodies conjugated to 12-nm gold particles (Methods). Electron micrographs visualize Cmu1-HA inside fungal hyphae, in the biotrophic interface and in the cytoplasm of infected maize cells. Leaves infected with SG200 and SG200 P_{cmu1}GFP-HA served as negative controls (Supplementary Figs 8 and 9). fc, fungal cytosol; fcw, fungal cell wall; bi, biotrophic interphase; pc, plant cytosol; pcw, plant cell wall; ppm, plant plasma membrane. Scale bars, 1 μ m. **B**, Electron micrographs of immunogold-labelled sections were analysed for the spatial distribution of gold labels in SG200 Δ cmu1-cmu1-HA (blue) and SG200 P_{cmu1}GFP-HA (red) infected tissue 3 days after infection. The total number of gold labels in each electron micrograph was set to 100% (see Methods for details). Error bars, s.d. of gold particles counted in three independent cross-sections. **C**, Confocal Z-stacks visualize spreading of Cmu1₂₂₋₂₉₀-mCherry to neighbouring tissue after biolistic transformation of maize leaves. White arrows indicate the originally transformed maize cells that carry a gold particle in their nucleus (**a**–**c**). Spreading of the fluorescent signal was observed in some cases for Cmu1-mCherry (**a**) and not in others (**b**, **c**). Yellow arrows mark Cmu1₂₂₋₂₉₀-mCherry signals in nuclei of neighbouring cells (**a**). Cmu1₂₂₋₂₉₀-mCherry spreading was never detected in transformed guard cells (**c**). Scale bars, 40 μ m.

ZmCm1 encodes a predicted plastidic isoform (Supplementary Fig. 16a) whereas ZmCm2 codes for a putative cytoplasmic enzyme (Supplementary Fig. 15b). Localization to the respective compartments was demonstrated by transient expression in maize leaves (Supplementary Fig. 16b, c). The observed compartmentalization mimics what has been described for the chorismate mutases of *Arabidopsis thaliana*¹⁸. Furthermore, as shown for the cytoplasmic isoform of chorismate mutase in *A. thaliana*¹⁹, ZmCm2 displayed enzymatic activity but no allosteric regulation *in vitro* (Supplementary Fig. 15c).

To demonstrate that the interaction between Cmu1 and the cytosolic maize chorismate mutase ZmCm2 can have functional consequences, we attempted to show that Cmu1 could alter ZmCm2 activity *in vitro*. As this was unsuccessful (Supplementary Fig. 15d), we next generated a loss of function allele of *cmu1* based on catalytically inactive forms of *S. cerevisiae* Aro7p (Supplementary Fig. 1a)^{20,21}. We reasoned that heterodimer formation between active and inactive monomers might interfere with chorismate mutase function. As expected, the *cmu1*_{R183A,K194A} allele was unable to complement the *aro7* deletion in *S. cerevisiae* (Fig. 1a). Surprisingly, when *cmu1*_{R183A,K194A} was introduced in single copy in either CL13Δ*cmu1* or SG200Δ*cmu1*, virulence was completely abolished (Supplementary Fig. 17). When *cmu1*_{R183A,K194A} was introduced in SG200 harbouring a functional *cmu1* allele, the mutated allele had a dominant effect that was copy number dependent (Supplementary Fig. 17). Confocal microscopy of infected leaf tissue revealed that the SG200Δ*cmu1*-*cmu1*_{R183A,K194A} strain could form appressoria (Supplementary Fig. 18a) but failed to colonize the plant (Supplementary Fig. 18b). In contrast to SG200 infections, plant cells infected with SG200Δ*cmu1*-*cmu1*_{R183A,K194A} were heavily stained by propidium iodide and displayed strong autofluorescence, probably because of the formation of phenolic compounds (Supplementary Fig. 18). This indicates that SG200Δ*cmu1*-*cmu1*_{R183A,K194A} elicits a strong plant defence response.

To exclude the possibility that the non-functional secreted chorismate mutase might interfere with the endogenous fungal shikimate pathway, we generated SG200Δ*cmu1* derivatives that express *cmu1*_{R183A,K194A} under control of a strong constitutive promoter (SG200Δ*cmu1*-P_{oter}*cmu1*_{R183A,K194A}-HA). Western blot analysis confirmed that the mutant protein was produced and secreted (Supplementary Fig. 19a). These strains did not show a growth phenotype on minimal media lacking aromatic amino acids, were morphologically indistinguishable from SG200 during growth in minimal media, and were unaltered in filamentous growth on charcoal media, a prerequisite for successful infection (Supplementary Fig. 19b–e). This illustrates that the secreted Cmu1_{R183A,K194A}-HA protein does not interfere with the activity of the cytoplasmic Aro7 protein in *U. maydis*. Cmu1_{R183A,K194A}-mCherry-HA accumulated around biotrophic hyphae like other secreted effectors¹³ (Supplementary Fig. 20b) but was unable to cause a dominant negative virulence phenotype when expressed in SG200Δ*cmu1* (Supplementary Fig. 20a). This suggests that plant uptake of Cmu1_{R183A,K194A}-mCherry-HA is necessary for the dominant effect on virulence, presumably because activity of ZmCm2 is affected. To obtain evidence that Cmu1_{R183A,K194A}-HA can reduce ZmCm2 activity, we first showed that ZmCm2 was able to interact with Cmu1_{R183A,K194A} and could complement the *aro7* mutation in *S. cerevisiae* (Supplementary Fig. 21). Next, *zmcm2* was co-expressed with *cmu1* or *cmu1*_{R183A,K194A} in the yeast *aro7* mutant strain. Although the co-expression of *zmcm2* and *cmu1* had no detectable effect on growth, co-expression of *zmcm2* and *cmu1*_{R183A,K194A} attenuated growth on plates lacking phenylalanine and tyrosine (Supplementary Fig. 21b). Therefore, the dominant negative effect on virulence elicited by the *cmu1*_{R183A,K194A}-HA allele is probably caused by interfering with the activity of cytosolic ZmCm2 through dimerization. This also implies that orphan cytosolic plant chorismate mutases might have an important regulatory function.

To obtain a comprehensive view on the metabolic changes in plants infected with CL13, CL13Δ*cmu1* and CL13Δ*cmu1*-*cmu1*-HA,

metabolome analyses were conducted 8 days after infection (Supplementary Figs 22 and 23 and Supplementary Table 2). Compared with mock-infected maize, plants infected with CL13 showed enhanced levels for phenylpropanoid and lignan biosynthesis products as well as for benzoxazinones, which derive from tryptophan (Supplementary Fig. 22 and Supplementary Table 3). For plants infected with CL13 and CL13Δ*cmu1*, the most notable differences concerned the phenylpropanoid pathway (Supplementary Fig. 22). Substances such as coumaroyl- and caffeoylquinic acid and syringic acid as well as lignan (like the syringaresinol-glucosides) were less abundant in tissue infected with CL13Δ*cmu1* than in plants infected with either CL13 or the complemented strain CL13Δ*cmu1*-*cmu1*-HA (Supplementary Fig. 22b and Supplementary Tables 2 and 3). In contrast, the amount of salicylic acid was at least ten times higher in plants infected with CL13Δ*cmu1* than those infected with the parental strain CL13 or CL13Δ*cmu1*-*cmu1*-HA, respectively (Fig. 1c). The amounts of the tryptophan-derived benzoxazinones were not significantly different in CL13Δ*cmu1* and CL13 infections (Supplementary Fig. 22), indicating that the pathway from chorismate to tryptophan through anthranilate synthase is unaffected by Cmu1 activity. The underlying mechanism for this differential effect awaits further study. Our results support a situation in which Cmu1 channels chorismate into the phenylpropanoid pathway and prevents its flow into the salicylic acid biosynthesis branch.

To elucidate the biological significance of the elevated salicylic acid levels in CL13Δ*cmu1* infections, maize seedlings were treated locally with 4 mM salicylic acid before infection or co-infiltrated during the infection with CL13. This concentration was chosen on the basis of total salicylic acid levels determined in CL13Δ*cmu1*-infected plants. Both treatments led to a reduction in virulence comparable to CL13Δ*cmu1* infections (Supplementary Fig. 24), which illustrates that salicylic acid enhances resistance of maize towards *U. maydis*. The data imply that the observed decrease in virulence for CL13Δ*cmu1* could be a direct consequence of its inability to interfere with pathogen-induced salicylic acid biosynthesis of the host plant.

Our findings provide new insights into a process that aids *U. maydis* during colonization of maize plants. It relies on the secretion of a chorismate mutase which enters plant cells by an unknown mechanism and redirects the metabolome in favour of the parasite.

We propose that the translocated fungal enzyme acts in conjunction with ZmCm2 in the plant cytosol by increasing the flow of chorismate from the plastid to the cytosol and in turn lowering the available substrate for salicylic acid biosynthesis in plastids (Fig. 3). However,

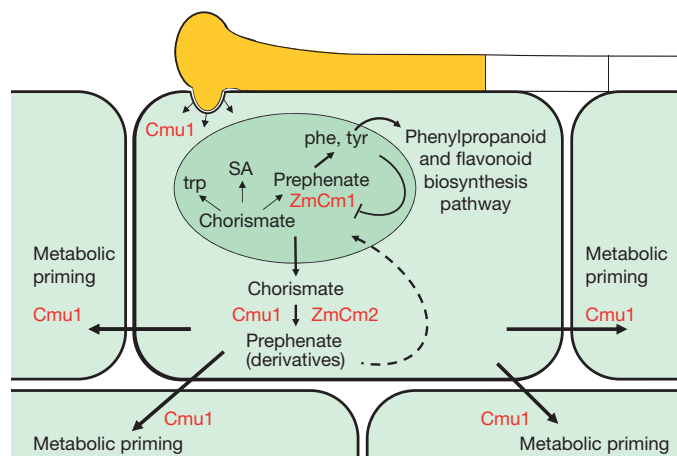


Figure 3 | Model of Cmu1-mediated metabolic priming in infected maize tissue. An infecting fungal hyphae is depicted in yellow. Maize cells are shown in mint green, the plastid is depicted in darker green. The dotted line indicates that prephenate or prephenate derivatives might be re-imported into the plastid or have regulatory capacity to feedback on plastidic synthesis of phenylalanine and tyrosine or derived phenolic compounds. SA, salicylic acid. For details, see text.

the introduction of a deregulated chorismate mutase into the host plant cytosol alone cannot explain all the metabolic changes observed in the *U. maydis* infected tissue (Supplementary Table 2 and Supplementary Fig. 18)^{2,22}. Thus, in line with its modest effects on virulence, we consider Cmu1 to be one component of a cocktail of effectors shaping the host metabolome. In this context it might not be coincidence that *U. maydis* has genes for two potential salicylate hydroxylases. In addition, organ-specific functions as described for several other *U. maydis* effectors cannot be excluded¹⁴.

The suppression of salicylic acid levels is likely to be particularly important for biotrophic pathogens and symbionts²³. In line with this we found genes encoding secreted chorismate mutases in many genomes of eukaryotic biotrophic plant pathogens and symbionts and several hemibiotrophic plant pathogens but only rarely in necrotrophic plant pathogens and fungal saprophytes (Supplementary Table 4). Recent findings indicate that the secreted chorismate mutase in the fungus *Sclerotinia sclerotiorum* might also represent a virulence factor (M. Dickman, personal communication). Metabolic priming by secreted chorismate mutases might thus emerge as a common strategy for host manipulation.

METHODS SUMMARY

The Methods section provides detailed information about all experimental procedures, including the following: (1) tables with details on oligonucleotides, plasmids, *U. maydis* and *S. cerevisiae* strains used or generated in this study; (2) details on the cloning strategies; (3) description of *U. maydis* mutant generation and their subsequent analysis; (4) links to bioinformatic tools applied in this study; (5) details for conducting quantitative real-time PCR analyses; (6) description of the yeast complementation assays; (7) description of yeast protein interaction assays; (8) details for conducting chorismate mutase activity assays; (9) the method to demonstrate protein secretion in *U. maydis*; (10) details on the transient expression in *Zea mays*; (11) confocal and electron microscopy methods; (12) details for metabolome and hormone analyses; and (13) protocol for the isolation and mass spectrometric analysis of secreted proteins of apoplastic fluids.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 13 February; accepted 12 August 2011.

Published online 5 October 2011.

- Doehlemann, G. *et al.* Pep1, a secreted effector protein of *Ustilago maydis*, is required for successful invasion of plant cells. *PLoS Pathog.* **5**, e1000290 (2009).
- Doehlemann, G. *et al.* Reprogramming a maize plant: transcriptional and metabolic changes induced by the fungal biotroph *Ustilago maydis*. *Plant J.* **56**, 181–195 (2008).
- Kamper, J. *et al.* Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**, 97–101 (2006).
- Mueller, O. *et al.* The secretome of the maize pathogen *Ustilago maydis*. *Fungal Genet. Biol.* **45** (suppl. 1), S63–S70 (2008).
- Bekal, S., Niblack, T. L. & Lambert, K. N. A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence. *Mol. Plant Microbe Interact.* **16**, 439–446 (2003).
- Doyle, E. A. & Lambert, K. N. *Meloidogyne javanica* chorismate mutase 1 alters plant cell development. *Mol. Plant Microbe Interact.* **16**, 123–131 (2003).
- Guermeur, Y., Geourjon, C., Gallinari, P. & Deleage, G. Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics* **15**, 413–421 (1999).

- Sasso, S., Ramakrishnan, C., Gamper, M., Hilvert, D. & Kast, P. Characterization of the secreted chorismate mutase from the pathogen *Mycobacterium tuberculosis*. *FEBS J.* **272**, 375–389 (2005).
- Eberhard, J., Raesecke, H.-R., Schmid, J. & Amrhein, N. Cloning and expression in yeast of a higher plant chorismate mutase. Molecular cloning, sequencing of the cDNA and characterization of the *Arabidopsis thaliana* enzyme expressed in yeast. *FEBS Lett.* **334**, 233–236 (1993).
- Krappmann, S. *et al.* The *aroC* gene of *Aspergillus nidulans* codes for a monofunctional, allosterically regulated chorismate mutase. *J. Biol. Chem.* **274**, 22275–22282 (1999).
- Mobley, E. M., Kunkel, B. N. & Keith, B. Identification, characterization and comparative analysis of a novel chorismate mutase gene in *Arabidopsis thaliana*. *Gene* **240**, 115–123 (1999).
- Basse, C. W., Kolb, S. & Kahmann, R. A maize-specifically expressed gene cluster in *Ustilago maydis*. *Mol. Microbiol.* **43**, 75–93 (2002).
- Doehlemann, G., Reissmann, S., Assmann, D., Fleckenstein, M. & Kahmann, R. Two linked genes encoding a secreted effector and a membrane protein are essential for *Ustilago maydis*-induced tumour formation. *Mol. Microbiol.* **81**, 751–766 (2011).
- Skibbe, D. S., Doehlemann, G., Fernandes, J. & Walbot, V. Maize tumors caused by *Ustilago maydis* require organ-specific genes in host and pathogen. *Science* **328**, 89–92 (2010).
- Bölker, M., Genin, S., Lehmle, C. & Kahmann, R. Genetic regulation of mating and dimorphism in *Ustilago maydis*. *Can. J. Bot.* **73**, 320–325 (1995).
- Di Stasio, M., Brefort, T., Mendoza-Mendoza, A., Munch, K. & Kahmann, R. The dual specificity phosphatase Rok1 negatively regulates mating and pathogenicity in *Ustilago maydis*. *Mol. Microbiol.* **73**, 73–88 (2009).
- Wille, A. C. & Lucas, W. J. Ultrastructural and histochemical studies on guard cells. *Planta* **160**, 129–142 (1984).
- Mobley, E. M., Kunkel, B. N. & Keith, B. Identification, characterization and comparative analysis of a novel chorismate mutase gene in *Arabidopsis thaliana*. *Gene* **240**, 115–123 (1999).
- Eberhard, J. *et al.* Cytosolic and plastidic chorismate mutase isozymes from *Arabidopsis thaliana*: molecular characterization and enzymatic properties. *Plant J.* **10**, 815–821 (1996).
- Schnappauf, G., Lipscomb, W. N. & Braus, G. H. Separation of inhibition and activation of the allosteric yeast chorismate mutase. *Proc. Natl Acad. Sci. USA* **95**, 2868–2873 (1998).
- Schnappauf, G., Strater, N., Lipscomb, W. N. & Braus, G. H. A glutamate residue in the catalytic center of the yeast chorismate mutase restricts enzyme activity to acidic conditions. *Proc. Natl Acad. Sci. USA* **94**, 8491–8496 (1997).
- Horst, R. J. *et al.* *Ustilago maydis* infection strongly alters organic nitrogen allocation in maize and stimulates productivity of systemic source leaves. *Plant Physiol.* **152**, 293–308 (2010).
- Glazebrook, J. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* **43**, 205–227 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are thankful to N. Amrhein and H.-U. Mösch for their comments on the manuscript. We thank B. Valent and C. H. Khang for alerting us to the fact that *Magnaporthe grisea* possesses a secreted chorismate mutase, and are grateful to M. Dickman for allowing us to cite his unpublished results. We thank T. Brefort, E. Mörschel, A. Kaever and M. Landesfeind for experimental support. We acknowledge advice by P. Kast, thank P. Schulze-Lefert for the Gateway-compatible plant transformation vectors, and D. Sicker for providing DIBOA and DIMBOA standards. We acknowledge technical assistance by R. Wissel, S. Löser, D. Vogel, F. Rath, G. Sowa, K. Bolte, M. Johannsen and P. Meyer. Our work was supported through DFG project DJ64/1-1, the collaborative research Center SFB593, and the LOEWE program of the State of Hesse.

Author Contributions A.D., K.S., F.R., V.V., J.K., S.O., T.T., K.F., P.M., Y.-D.S., H.S., A.G. and B.M. designed and performed the wet bench experiments. All authors contributed to data analysis. R.K., A.D. and K.S. wrote the manuscript with input from all co-authors. R.K. directed the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.K. (kahmann@mpi-marburg.mpg.de).

METHODS

Generation of plasmids. Standard molecular cloning strategies and techniques were applied in this study²⁴. Most of the constructs were generated using Gateway technology (Invitrogen) after an insertion of the gene of interest into the pEntry4 vector (Invitrogen) using NcoI and NotI restriction sites. Primers used in this study are described in Supplementary Table 5. Plasmids that were generated in this study are listed in Supplementary Table 6.

Mutant generation and analysis. All *U. maydis* strains (Supplementary Table 7) were generated by gene replacement with PCR-generated constructs or by insertion of p123 derivatives into the *ip* locus as described²⁵ (Supplementary Table 7). At least three independent mutants were repeatedly tested for virulence on 7-day-old maize seedlings and disease was scored 12 days after infection following described protocols³. The widely used solopathogenic haploid strains SG200 and CL13 differ in virulence owing to the presence of autocrine pheromone signalling in SG200 and its absence in CL13, respectively¹⁶. Compared with the naturally occurring dikaryon, both strains show reduced virulence. Typical symptoms caused by CL13 are depicted in Supplementary Fig. 3a.

Bioinformatic analyses. Signal peptide prediction was performed with the program SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>). Chloroplast transit peptides were predicted with the program ChloroP (<http://www.cbs.dtu.dk/services/ChloroP/>). Sequence alignments were generated using CloneManager Suite 9.0 (www.scied.com). Hierarchical neural network was applied for prediction of the Cmu1 secondary structure at the Network Protein Sequence Analysis server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_nn.html). Domain analyses were performed with Smart and InterPro (<http://smart.embl-heidelberg.de/>; <http://www.ebi.ac.uk/Tools/InterProScan/>).

Quantitative real-time PCR. RNA was extracted from sporidia grown in axenic culture as well as from infected maize plants at the indicated time points with the TRIzol method (Invitrogen), treated with DNase (Ambion) and subsequently used for cDNA synthesis. Quantitative real-time PCR reactions were conducted as described earlier². All reactions were performed at least in biological triplicates. Relative *cmu1* expression levels were calculated in relation to the values obtained for the constitutively expressed peptidyl-prolyl *cis-trans* isomerase gene (*ppi*) of *U. maydis*²⁶ (Supplementary Table 5).

Yeast complementation assay. Yeast strain Y054679 lacking the *ARO7* gene was transformed with the corresponding pYES and pGad derivatives (Supplementary Table 6) using standard protocols (Clontech) and tested for growth on medium lacking phenylalanine and tryptophan as described previously¹⁹. A compilation of all *S. cerevisiae* strains used in this study is provided in Supplementary Table 8.

Yeast protein interaction assay. The genes encoding the proteins tested for interaction were cloned into pGBKT7 or pGADT7 vectors (Clontech; Supplementary Table 6), generating in-frame fusions with a gene encoding the yeast *GAL4* binding and activation domain, respectively. Interaction was tested in *S. cerevisiae* AH109 (Clontech). Growth controls were performed on selective dropout media (SD) plates lacking only tryptophan and leucine to select for cells containing the correct plasmids. Protein interactions were assayed on high-stringency SD plates additionally lacking adenine and histidine.

Chorismate mutase activity assays. A glutathione S-transferase (GST)–Cmu1_{22–290}–HA fusion protein was produced in *Escherichia coli* BL21 containing plasmid pRset-cmu_{22–290}–HA (Supplementary Table 6) and enriched by glutathione-affinity purification (GE Healthcare). Also, GST–ZmCm2 fusion protein and derivatives of Cmu1 for the enzyme assay were made accordingly. After removal of the GST moiety using PreScission protease (GE Healthcare), chorismate mutase activity assays were performed²⁷. After acidic conversion of prephenate to phenylpyruvate the reaction was basified and extinction at $\lambda = 320$ nm was measured. The increase in extinction was plotted against time (in minutes) to visualize the formation of phenylpyruvate. Error bars represent s.d. from three technical replicates. Purified GST protein was used as a negative control and respective values were subtracted from those obtained with Cmu1_{22–290}–HA.

Demonstration of Cmu1 secretion. *U. maydis* strain AB33 P_{otef}cmu1–HA was generated by insertion of plasmid p123_{otef}:um05731–HA into the *ip* locus of AB33²⁸ (Supplementary Table 7). To analyse Cmu1–HA secretion, material was collected 6 h after induction of filamentous growth in medium containing nitrate²⁸. Protein extracts of filamentous cells and culture supernatants (after precipitation with trichloroacetic acid) were subjected to western blot analysis with mouse-anti HA (Sigma) and mouse anti- α -tubulin antibodies (Oncogene).

Biolistic transformation of *Z. mays*. For biolistic transformation²⁹ of 7- to 10-day-old maize leaves, 1.6- μ m gold particles were coated with plasmid DNA coding for the indicated genes driven by the CaMVS35 promoter (Supplementary Table 6). Bombardment was performed using a PDS-1000/HeTM instrument (BioRad) at 900 p.s.i. in a 27 Hg vacuum. Fluorescence was observed by confocal microscopy 2 days after transformation.

Confocal and electron microscopy. Confocal microscopy was performed with a LeicaSP5 confocal microscope as described³⁰. Wheat germ agglutinin/Alexa Fluor 488 and propidium iodide stains were performed as reported³⁰. Autofluorescence was detected at $\lambda = 415$ –460 nm.

For immunogold labelling, infected leaf parts were cryofixed by high-pressure freezing (Bal-Tec HPM 010), freeze-substituted in 0.5% glutaraldehyde in acetone (containing 2% H₂O), infiltrated with Lowicryl HM20 and ultraviolet-polymerized at -40 °C. Ultrathin sections were labelled for HA epitope detection using mouse anti-HA (Sigma H9658) and donkey anti-mouse 12-nm gold antibodies (Jackson 715-205-150) and imaged in a Philips CM10 electron microscope at 60 kV.

The distribution of gold particles was determined semi-quantitatively as described³¹. Micrographs of sectioned *Z. mays* samples from infections with SG200 P_{cmu1}Cmu1–HA and SG200 P_{cmu1}GFP–HA were selected and in each case three different hyphae were chosen. Gold particles on each of the micrographs were then counted and assigned to the plant cell cytosol, the biotrophic interface or the fungal cytosol. The proportional distribution in these compartments was then calculated as a percentage and the s.d. was calculated from the three different data sets for each sample.

Metabolome analyses. For metabolite fingerprinting a section of the third leaf between 1 and 3 cm below the injection holes was excised 8 days after syringe infection with *U. maydis* strains or water (mock control), respectively. For each replicate, 30–40 leaf sections were pooled. Plant material was homogenized under liquid nitrogen. Two or three biological replicates of control leaves and infected leaves (80 mg each) were extracted with methyl-*tert*-butylether/methanol³². The polar phase was dried under a nitrogen stream and the extracted metabolites resolved in 10 μ l of methanol, 10 μ l acetonitrile and 120 μ l water. The metabolite analysis was performed by ultra-performance liquid chromatography (UPLC, ACQUITY UPLC System, Waters Corporation) coupled with an orthogonal time-of-flight mass spectrometer (TOF-MS, LCT Premier, Waters Corporation). For LC an ACQUITY UPLC BEH SHIELD RP18 column (1 mm \times 100 mm, 1.7 μ m particle size, Waters Corporation) was used at a temperature of 40 °C, a flow rate of 0.2 ml min⁻¹ and with the following gradient for the analysis of the polar phase: 0–0.5 min 10% B, 0.5–3 min from 10% B to 28% B, 3–8 min from 28% B to 95.5% B, 8–10 min 95.5% B and 10–14 min 10% B (solvent system A: water/formic acid (100:0.1, v/v); B: acetonitrile/formic acid (100:0.1, v/v)). The TOF-MS was operated in negative as well as positive electrospray ionization mode in W optics with a mass resolution larger than 10,000. Data were acquired by MassLynx software (Waters Corporation) in centroided format over a mass range of m/z 85–1,200 with scan duration of 0.5 s and an interscan delay of 0.1 s. The capillary and the cone voltage were maintained at 2,700 V and 30 V and the desolvation and source temperature at 350 °C and 80 °C, respectively. Nitrogen was used as cone (30 l h⁻¹) and desolvation gas (800 l h⁻¹). For accurate mass measurement, the TOF-MS was calibrated with phosphoric acid 0.01% (v/v) in acetonitrile/water (50:50, v/v) and the dynamic range enhancement mode was used for data recording. All analyses were monitored by using leucine-enkephalin ($[M + H]^+$ 556.2771 or $[M - H]^-$ 554.2615 as well as its ¹³C isotopomer $[M + H]^+$ 557.2803 or $[M - H]^-$ 555.2615, Sigma-Aldrich) as lock spray reference compound at a concentration of 0.5 μ g ml⁻¹ in acetonitrile/water (50:50, v/v) and a flow rate of 30 μ l min⁻¹. The raw mass spectrometry data of all samples were processed using the MarkerLynx Application Manager for MassLynx software (Waters Corporation), resulting in two data sets.

The toolbox MarVis (<http://marvis.gobics.de33>) was used for ranking, filtering, adduct correcting and combining the data as well as for clustering and visualization, respectively. An analysis of variance test was applied to extract a subset of high-quality marker candidates with a p value less than 1×10^{-5} . The filtered data sets were adduct corrected according to the following rules: $[M + H]^+$, $[M + Na]^+$, $[M + NH_4]^+$ for the positive and $[M - H]^-$, $[M + CH_2O_2 - H]^-$, $[M + CH_2O_2 + Na - 2H]^-$ for the negative ionization mode. The combined data led to an overall data set of 810 marker candidates (Supplementary Table 2), which were used for clustering and visualization by means of one-dimensional self-organizing-maps and for database search.

The identity of selected markers was confirmed by MS² fragment information³⁴, co-elution with identical standards or exact mass measurement (Supplementary Table 3).

Salicylic acid measurements. For metabolite fingerprinting, a section of the third leaf between 1 and 3 cm below the injection holes was excised 8 days after syringe infection with *U. maydis* strains or water (mock control), respectively. For each replicate, 30–40 leaf sections were pooled.

Total salicylic acid was extracted³⁵ and identified by co-elution with an authentic standard using liquid chromatography–mass spectrometry.

Proteome analysis of apoplastic fluids. To extract apoplastic fluids, maize seedlings were infected with a mixture of FB1 and FB2 (ref. 36). Two, four and six days after infection, infected areas were excised and apoplastic fluid was

- collected³⁷. After precipitation with trichloroacetic acid, proteins were separated by 12% SDS–polyacrylamide gel electrophoresis, digested in gel³⁸ after subdividing each lane into 11 equal parts and run on an Agilent 1100 nano-HPLC system (75- μ m C18 column, 100-min gradients), coupled to an LTQ-FT mass spectrometer (Thermo Scientific). The ‘Top-3-SIM’ acquisition method was used, as described³⁹. Spectra were processed by MSQuant⁴⁰ and searched using Mascot against a decoy *Zea/Ustilago* protein database. Mass tolerance for the precursor ion was in all cases 5 p.p.m, and for fragment ions 0.5 Da; full trypsin specificity was required and two missed cleavages were allowed. The mean measurement mass deviation of precursor (peptide) ions was 0.96 p.p.m. with a standard deviation of 0.82.
24. Sambrook, J. F. E. & Maniatis, T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, 1200 (1989).
 25. Loubradou, G., Brachmann, A., Feldbrugge, M. & Kahmann, R. A homologue of the transcriptional repressor Ssn6p antagonizes cAMP signalling in *Ustilago maydis*. *Mol. Microbiol.* **40**, 719–730 (2001).
 26. Bohlmann, R. *Isolierung und Charakterisierung von filamentspezifisch exprimierten Genen aus Ustilago maydis*. PhD thesis, Ludwig-Maximilian-Univ. München (1996).
 27. Gilchrist, D. G. C. & Conelly, J. A. Chorismate mutase from mung bean and sorghum. *Methods Enzymol.* **142**, 450–463 (1987).
 28. Brachmann, A., Weinzierl, G., Kamper, J. & Kahmann, R. Identification of genes in the bW/bE regulatory cascade in *Ustilago maydis*. *Mol. Microbiol.* **42**, 1047–1063 (2001).
 29. Ueki, S., Lacroix, B., Krichevsky, A., Lazarowitz, S. G. & Citovsky, V. Functional transient genetic transformation of *Arabidopsis* leaves by biolistic bombardment. *Nature Protocols* **4**, 71–77 (2008).
 30. Doehlemann, G. *et al.* Establishment of compatibility in the *Ustilago maydis*/maize pathosystem. *J. Plant Physiol.* **165**, 29–40 (2008).
 31. Shimada, Y., Ichinose, S., Sadr, A., Burrow, M. F. & Tagami, J. Localization of matrix metalloproteinases (MMPs-2, 8, 9 and 20) in normal and carious dentine. *Aust. Dent. J.* **54**, 347–354 (2009).
 32. Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J. Lipid Res.* **49**, 1137–1146 (2008).
 33. Kaever, A. *et al.* MarVis: a tool for clustering and visualization of metabolic biomarkers. *BMC Bioinformatics* **10**, 92 (2009).
 34. Pommerrenig, B. *et al.* Phloem-specific expression of yang cycle genes and identification of novel yang cycle enzymes in plantago and *Arabidopsis*. *Plant Cell* **23**, 1904–1919 (2011).
 35. Naranjo, M. A. *et al.* Lithium treatment induces a hypersensitive-like response in tobacco. *Planta* **217**, 417–424 (2003).
 36. Banuett, F. & Herskowitz, I. Different alleles of *Ustilago maydis* are necessary for maintenance of filamentous growth but not for meiosis. *Proc. Natl Acad. Sci. USA* **86**, 5878–5882 (1989).
 37. De Wit, P. J. G. M. & Spikman, G. Evidence for the occurrence of race and cultivar-specific elicitors of necrosis in intercellular fluids of compatible interactions of *Cladosporium fulvum* and tomato. *Physiol. Plant Pathol.* **21**, 1–11 (1982).
 38. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protoc.* **1**, 2856–2860 (2006).
 39. Olsen, J. V. & Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci USA* **101**, 13417–13422 (2004).
 40. Mortensen, P. *et al.* MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. *J. Proteome Res.* **9**, 393–403 (2010).

DNA stretching by bacterial initiators promotes replication origin opening

Karl E. Duderstadt¹, Kevin Chuang² & James M. Berger^{1,2}

Many replication initiators form higher-order oligomers that process host replication origins to promote replisome formation. In addition to dedicated duplex-DNA-binding domains, cellular initiators possess AAA+ (ATPases associated with various cellular activities) elements that drive functions ranging from protein assembly to origin recognition. In bacteria, the AAA+ domain of the initiator DnaA has been proposed to assist in single-stranded DNA formation during origin melting. Here we show crystallographically and in solution that the ATP-dependent assembly of *Aquifex aeolicus* DnaA into a spiral oligomer creates a continuous surface that allows successive AAA+ domains to bind and extend single-stranded DNA segments. The mechanism of binding is unexpectedly similar to that of RecA, a homologous recombination factor, but it differs in that DnaA promotes a nucleic acid conformation that prevents pairing of a complementary strand. These findings, combined with strand-displacement assays, indicate that DnaA opens replication origins by a direct ATP-dependent stretching mechanism. Comparative studies reveal notable commonalities between the approach used by DnaA to engage DNA substrates and other, nucleic-acid-dependent, AAA+ systems.

All organisms depend on ring- and spiral-shaped ATPase assemblies to carry out essential processes ranging from proteolysis and membrane trafficking, to signalling events and nucleic acid transactions. DNA replication onset in cells reflects one such process, using ATP-dependent initiation factors to coordinate replisome assembly^{1,2}. Replication initiators of eukaryotes and prokaryotes contain AAA+-family ATPase domains, the activity of which is augmented by duplex-DNA-binding domains and specialized protein-protein interaction elements that assist with origin recognition and recruit specific replication factors^{3,4}. Although all AAA+ enzymes share a common structural core with RecA-type ATPases, together forming the additional strand catalytic glutamate (ASCE) supergroup of P-loop NTPases⁵, the molecular logic that allows a common nucleotidyl-hydrolase module to control the disparate activities of replication initiators, and ASCE proteins in general, is not understood.

In bacteria, replication initiation relies on the DnaA protein^{6–8}. In *Escherichia coli*, multiple DnaA molecules bind to the replication origin, *oriC*, through several duplex DNA-binding sites, forming a large nucleoprotein complex in the presence of ATP^{9–11}. With the aid of appropriate architectural proteins (such as integration host factor) and negatively supercoiled DNA, this complex subsequently melts an (A+T)-rich, DNA-unwinding element (DUE) located adjacent to the duplex DnaA binding sites^{12,13}. ATP also activates a secondary DNA-binding site within DnaA, thought to reside within the AAA+ domain, which engages single-stranded regions of the DUE to form a stable open complex^{12,14–16}. DnaA then collaborates with the bacterial helicase loader (DnaC in *E. coli*) to recruit two hexamers of the DnaB helicase to the origin and promote replisome assembly^{17–19}.

Although most AAA+ enzymes form closed-ring assemblies^{20,21}, structural studies have indicated that initiators and polymerase clamp-loaders form open-ring structures^{14,22–24}. Among initiator/loader systems, DnaA is particularly unusual in that it has been seen to oligomerize into a right-handed, spiral filament¹⁴. Two models have been proposed to explain how this structure might aid origin melting (Supplementary Fig. 1). In one, the wrapping of duplex DNA about a

DnaA superhelix would constrain a positive supercoil, generating compensatory negative writhe that could aid opening of the neighbouring DUE. In the other, the wrapped DnaA–DNA complex would serve as a nucleation centre, allowing DnaA protomers to engage directly and melt the DUE, possibly through the initiator's ATPase elements. Thus far, experimental evidence has supported both models^{9,14–16,25}, leaving open the question as to how DnaA catalyses origin melting. The relationship of this mechanism to other initiation systems, or to AAA+/ASCE proteins overall, is also unclear.

A DnaA–ssDNA crystal structure

To examine these issues, we set out to determine the structure of DnaA bound to single-stranded DNA (ssDNA). Using a truncation of *Aquifex aeolicus* DnaA consisting of the AAA+ and duplex-DNA-binding domains (which, like its *E. coli* counterpart^{16,17}, is active for both ATP-stimulated assembly and ssDNA binding²⁵), we first grew DNA-free crystals in the presence of Mg²⁺ and the non-hydrolysable ATP mimic AMPPCP¹⁴. DNA substrates were then soaked into these crystals under low-salt conditions (Methods). Data collection and phasing by molecular replacement revealed four DnaA protomers per asymmetric unit, arranged in a spiral configuration that propagates into a continuous protein helix by the action of crystal-symmetry elements (Fig. 1a, b), along with bound ssDNA. Of the multiple substrates screened (Methods), dA₁₂ yielded the highest-quality density (Supplementary Fig. 2a, b), and served as the best target for model building and refinement. The final structure, containing a DnaA:AMPPCP:Mg²⁺:dA₁₂ stoichiometry of 4:4:4:1, was refined to an $R_{\text{work}}/R_{\text{free}}$ of 24.9/26.8% at 3.35 Å resolution (Supplementary Table 1).

DnaA–ssDNA interactions

The overall arrangement of DnaA subunits in the helical assembly is highly similar to a DNA-free form reported previously (0.7 Å root mean squared deviation between all C α positions)¹⁴. AMPPCP•Mg²⁺ binds at the interface between neighbouring subunits, with the γ -phosphate of AMPPCP coordinated by catalytic amino acids from pairs of adjoining

¹Biophysics Graduate Group, University of California, Berkeley, Berkeley, California 94720, USA. ²Department of Molecular and Cell Biology, California Institute for Quantitative Biology, University of California, Berkeley, Berkeley, California 94720, USA.

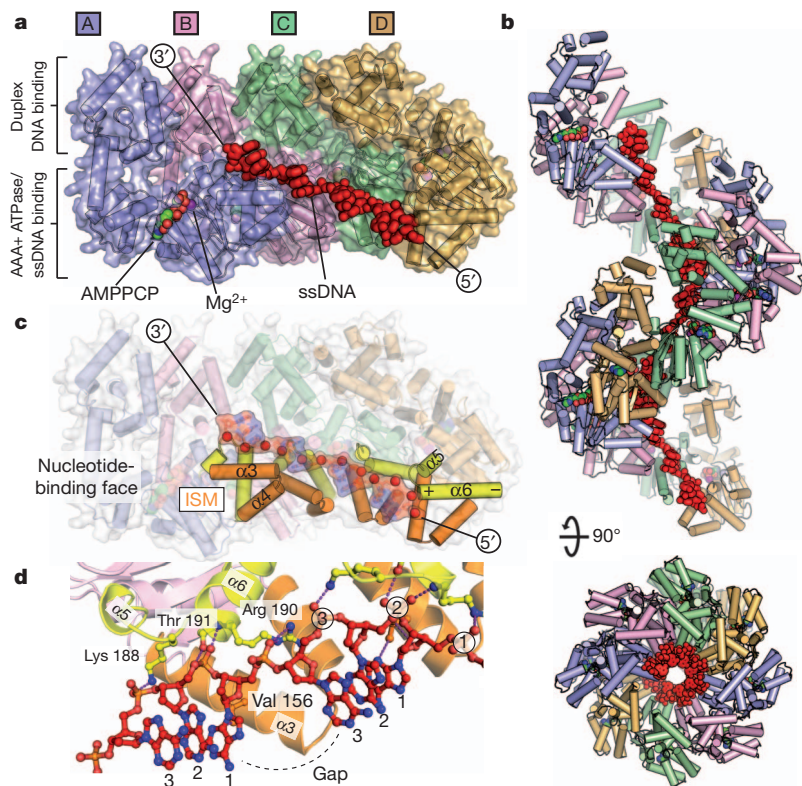


Figure 1 | The ATPase pore of assembled DnaA binds ssDNA. **a**, Side view of the asymmetric unit, with DnaA subunits differentially coloured. Single-stranded DNA is displayed as red sticks. AMPPCP and Mg^{2+} , bound to chain A, are shown as spheres coloured by element and in magenta, respectively; AMPPCP• Mg^{2+} bound to chains B–D is occluded in this view. **b**, Side and top views of oligomerized DnaA, reconstructed through crystal packing, showing 12 DnaA subunits and 3 strands of ssDNA. Colouring as in panel **a**. **c**, Side view of the DnaA tetramer with helices $\alpha 3/\alpha 4$ and $\alpha 5/\alpha 6$ highlighted in orange and yellow, respectively. ISM, initiator-specific motif. Single-stranded DNA is shown as a transparent stick-and-surface representation coloured by element; phosphates are further highlighted as red spheres. **d**, Protein–DNA contacts. Protein chains B (left) and C (right) are displayed with the same colouring as in panel **c**. Single-stranded DNA is coloured by element.

AAA+ domains. Single-stranded DNA associates exclusively with the AAA+ elements of the initiator, with each protomer binding three nucleotides of the dA_{12} strand (Fig. 1a). Almost all contacts are made through the phosphodiester backbone, exposing the DNA bases to solvent. Each trinucleotide segment adopts a B-form DNA conformation (Supplementary Fig. 3), with the bases between consecutive segments separated by large (~ 10 Å) gaps that extend the substrate by $\sim 50\%$ (Supplementary Table 2 and Supplementary Information).

DnaA binds ssDNA using just two pairs of helices, $\alpha 3/\alpha 4$ and $\alpha 5/\alpha 6$, both of which line the central channel of the protein assembly (Fig. 1c). The geometry of these two elements creates a single conduit along the length of the DnaA superhelix that allows substrate to traverse consecutive DnaA protomers. Interestingly, helices $\alpha 3/\alpha 4$ also comprise the initiator-specific motif, which both promotes filament formation^{14,19} and distinguishes DnaA as a member of the initiator clade of the AAA+ superfamily^{26,27}.

DnaA uses a simple network of interactions to coordinate ssDNA. The initiator-specific motif forms a shelf for each trinucleotide, in which a conserved hydrophobic residue, Val 156, forms van der Waals contacts with the sugar and base of the first nucleotide in the triplet (Fig. 1d). The central phosphate of each trinucleotide is bound by the electropositive, amino-terminal helix dipole of $\alpha 6$ and hydrogen bonded by Thr 191 (Fig. 1c, d). These contacts are flanked by two positively charged residues, Arg 190 and Lys 188, which make salt-bridge interactions with the phosphates of nucleotides 1 and 3, respectively. Notably, mutant initiators containing substitutions in these observed DNA-binding residues show reduced affinity for ssDNA in solution (Supplementary Fig. 4), confirming that the crystals captured a physiologically meaningful initiator state. Moreover, mutations of the same positions in *E. coli* DnaA (amino acids Arg 245, Lys 243 and Val 211) also disrupt ssDNA binding and origin melting¹⁵. Thus, the ssDNA engagement strategy seen here seems to be conserved across bacterial species.

Structural similarities between DnaA and RecA

In considering the assembly patterns of oligomeric ATPases, we were struck by the similarity of DnaA to one system in particular: the

homologous recombination protein, RecA. Although the cellular functions of these two proteins are fundamentally different (catalysis of DNA strand-exchange reactions versus replication origin melting and coordination of replisome assembly), both RecA and DnaA are predicated upon an ASCE ATPase fold^{27,28}. Like DnaA, RecA (and its Rad51/RadA orthologues) forms a helical assembly that engages DNA with its pore regions^{28–32}. These shared physical properties led us to undertake a more detailed comparison of RecA and DnaA. Of the multiple models available, the structure of a RecA oligomer bound to ssDNA³³, representing the presynaptic complex formed during the initial stages of homologous recombination, is globally most similar to the DnaA state we observe (Fig. 2a, b). As with DnaA, RecA contacts DNA almost exclusively through the phosphodiester backbone, which sits in the interior of a positively charged filament pore. Each RecA protomer binds three nucleotides in a B-DNA conformation, with the base stacking between each triplet interrupted such that ssDNA is extended ~ 1.5 -fold compared to a B-form duplex (Fig. 2c).

RecA and DnaA also exhibit some interesting and significant differences. A visual examination of each triplet shows that RecA uses a more extensive network of contacts for engaging ssDNA than does DnaA (Fig. 2d, e), burying twice as much surface area per triplet (318 Å² and 639 Å² for DnaA and RecA, respectively). This difference derives largely from an additional β -hairpin in RecA that fills the gap between each triplet and reinforces each three-base stack³³. Moreover, whereas two of the three nucleotides within each RecA triplet (positions 1 and 2) align well with those seen in DnaA, position 3 of the DnaA trinucleotide rotates away from the pore axis by $\sim 50^\circ$ (Fig. 2f). This difference skews consecutive DnaA triplets away from one another, disrupting the formation of a smoothly spiral arrangement as seen in RecA (Fig. 2c).

DNA extension is ATP- and assembly-dependent

The ability of RecA to stretch DNA to the extent observed crystallographically has been amply substantiated by various methodologies^{34–36}. Using these efforts as a guide, we set out to determine whether the DNA conformation that we observe bound to DnaA accurately represents the state of the substrate in solution. To accomplish this, we turned to a

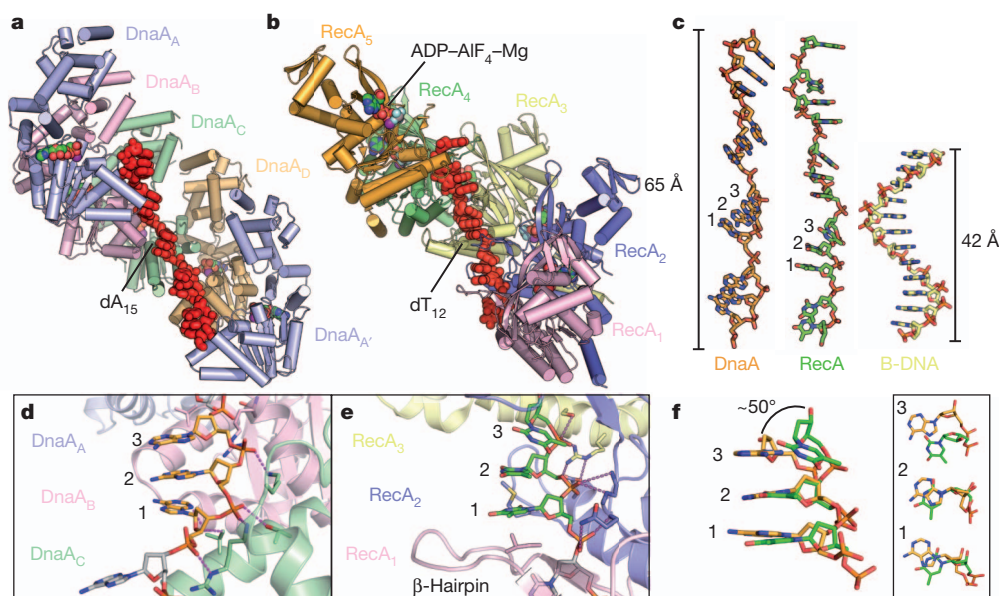


Figure 2 | DnaA engages ssDNA in a manner similar to RecA. **a**, View of a DnaA-AMPPCP-ssDNA pentamer (consisting of one full tetramer, as well as chain A (DnaA_{A'}) and its associated triplet from the adjacent asymmetric unit). AMPPCP•Mg²⁺ is shown as spheres coloured by atom; ssDNA as red sticks. **b**, View of a RecA-ADP-AlF₄-ssDNA pentamer (Protein Data Bank (PDB) accession 3CMW)³³. ADP•AlF₄•Mg²⁺ is shown as spheres coloured by atom;

ssDNA as red sticks. **c**, Comparison of ssDNA bound to DnaA (orange), RecA (green) and a strand of B-DNA (yellow). **d**, Close-up view of triplet bound to DnaA (chain C) with magenta dashed lines indicating key contacts. **e**, Close-up view of triplet bound to RecA (protomer 2) with magenta dashed lines indicating key contacts. **f**, Side (left) and top (right) views of the triplets displayed in **d** and **e** aligned with each other.

bulk-phase fluorescence resonance energy transfer (FRET)-based ssDNA extension assay analogous to single-molecule approaches applied to RecA³⁷. Using a poly-thymine DNA labelled with Cy3 and Cy5 (FR-dT₂₁) (Supplementary Table 3), we monitored changes in the length of ssDNA resulting from DnaA binding (Fig. 3a). Analogous studies were performed with RecA as a control. As both RecA and DnaA require ATP for formation of the oligomers observed in the structural models, we expected ATP-dependent extension to lead to

a loss of FRET signal. We tested for extension both in the presence of the ATP analogues ATPγS and ADP•BeF₃, to avoid complications that might arise from nucleotide hydrolysis, and in the presence of ADP, which is known to promote DnaA disassembly. Pronounced extension was observed only in the presence of the ATP analogues (Fig. 3b, c), and not with ADP. The lengths of ssDNA in the ATP-assembled states of both proteins, as calculated from the FRET data, were in close agreement with those observed in the crystal structures (Supplementary Table 6). Likewise, mutations in ssDNA-binding amino acids and residues required for DnaA assembly all significantly reduced ssDNA extension (Supplementary Fig. 6), demonstrating that this activity depends on substrate binding to the pore of an initiator oligomer that forms only when activated by ATP.

DnaA directly catalyses duplex melting

How replication origins are opened for replisome assembly is an important, unanswered question. Given the similarities between the ssDNA binding and extension activities of DnaA and RecA, we reasoned that the initiator might directly destabilize and disrupt DNA duplexes. This activity is a known property of RecA³⁸, albeit one that permits the recombinase to exchange DNA strands between target substrates actively^{30,33}.

To test this idea, we developed a DNA strand-displacement assay for DnaA. First, the initiator was incubated with a short duplex containing one fluorescently labelled strand. Unlabelled competitor strand was then added to capture any unwound species (Fig. 4a). Both ADP and ADP•BeF₃ were tested to determine whether initiator assembly affected the outcome of the experiment, as were DNAs of different lengths and stabilities. Analysis of the resultant products by gel electrophoresis shows that DnaA readily unwinds a 15mer duplex DNA of moderate stability ($T_m = 43^\circ\text{C}$) in the presence of the ATP mimic (Fig. 4b). By contrast, increasing the stability of the DNA substrate by ~30% (using a 20mer, $T_m = 55^\circ\text{C}$) weakens the unwinding activity of DnaA (Fig. 4b), while increasing DNA stability even further (30mer, $T_m = 62^\circ\text{C}$) abrogates melting completely (Supplementary Fig. 7a). Importantly, ADP did not support strand displacement, nor did ssDNA binding and DnaA assembly mutants (Supplementary Fig. 7b, c). These controls indicate that double-stranded-DNA melting is dependent not

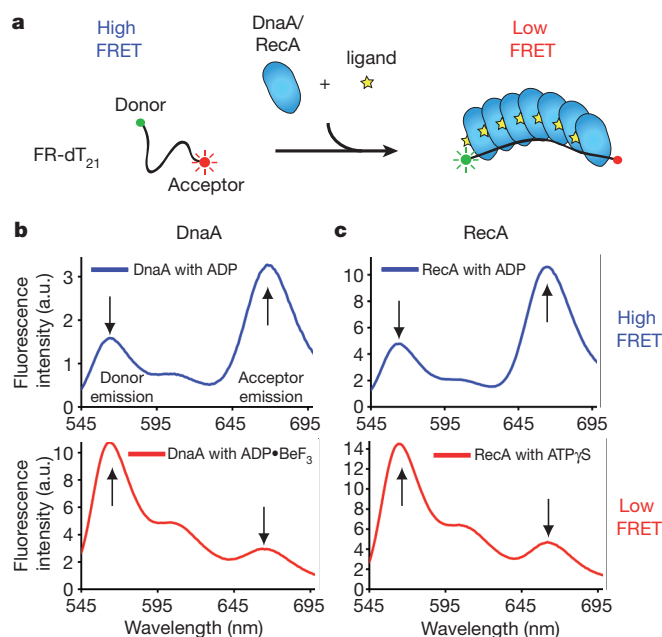


Figure 3 | DnaA extends ssDNA in solution. **a**, Cartoon of ssDNA extension assay. **b**, Emission scan (donor excitation) of FR-dT₂₁ in the presence of 10 μM DnaA with either ADP•BeF₃ (top) or ADP (bottom). **c**, Emission scan (donor excitation) of FR-dT₂₁ in the presence of 10 μM RecA with either ATPγS (top) or ADP (bottom). Reported transfer efficiencies and distances were calculated using donor emission as described in Methods. a.u., arbitrary units.

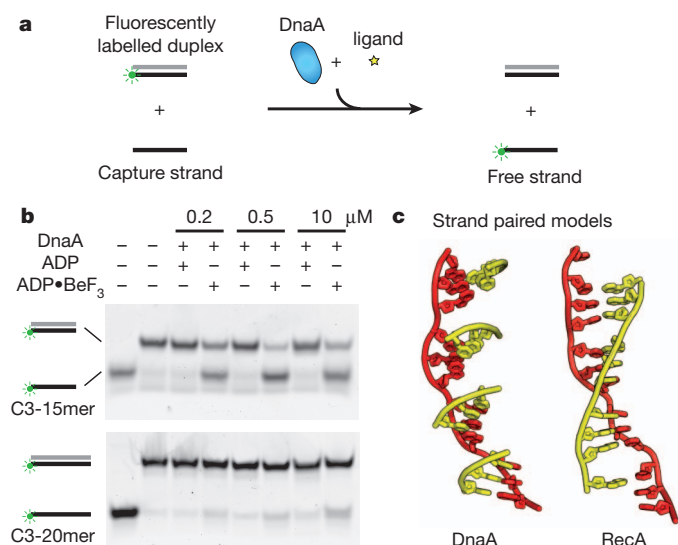


Figure 4 | DnaA directly melts duplex DNA. **a**, Schematic of strand displacement assay. The green circle represents the Cy3 fluorescent end-label used to follow the status of one DNA strand. Complementary strands of duplex substrates are coloured grey and black. **b**, Strand displacement assay conducted with 15mer and 20mer duplex substrates (C3-15mer and C3-20mer) in the presence and absence of different nucleotides. DnaA concentrations used are indicated above each lane. **c**, Left: cartoon model showing how complementary base triplets (yellow) would pair (in a B-DNA manner) with ssDNA bound to DnaA (red). The orientation of successive DnaA-bound triplets is such that it prevents the formation of a continuous base-paired strand favouring duplex separation. Right: same DNA view, but as seen in RecA, where triplets are oriented to allow pairing of an extended complementary strand to promote duplex formation and strand exchange (PDB accession 3CMX)³³.

only on formation of an assembled DnaA oligomer, but that the initiator is fine-tuned to specifically disrupt DNAs of modest stability.

One significant functional difference between RecA and DnaA is that the recombination protein can drive a true strand-exchange reaction; that is, in addition to displacing one strand of a duplex, RecA can also pair homologous ssDNA segments into a double-stranded molecule. By contrast, the function of DnaA is to separate double-stranded origin regions. Inspection of the RecA and DnaA complexes reveals a physical basis for these differing properties: in DnaA, successive trinucleotide elements are arranged in a state incompatible with the formation of a continuous duplex, whereas

ssDNA bound to RecA adopts a smoothly spiralled arrangement permitting the contiguous pairing of a complementary strand (Fig. 4c). This distinction arises primarily from the 50° rotation between the nucleotides at the third position of each triplet seen in the RecA and DnaA models (Fig. 2f). In DnaA, the orientation of this nucleotide appears to be stabilized by base stacking, whereas in RecA the β -hairpin insertion helps to sculpt the configuration of the DNA to create a contiguous base-pairing surface.

Implications for origin melting

Together, our findings present the strongest evidence yet that DnaA melts replication origins by directly assisting with the separation and sequestration of duplex DNA strands (Supplementary Fig. 1c). Notably, this activity does not contradict the demonstrated need for other factors capable of reshaping and/or destabilizing DNA (for example, integration host factor and negative supercoiling) during initiation^{12,13}. Rather, these elements probably help to promote DnaA assembly and prime the origin for melting by what otherwise would be an inefficient unwindase. In this view, the AAA+ domains of DnaA may first engage only one of the two strands of duplex DNA with their ssDNA binding elements (possibly at reported ssDNA or ATP-DnaA binding sites^{15,16}). In the presence of ATP, which triggers initiator assembly, subunit-subunit interactions would help to restructure the DNA backbone, stretching the contacted strand to facilitate melting. Re-annealing would be disfavoured by the non-contiguous arrangement of base triplets in the extended state (Fig. 4c). Future studies will be needed to define the specific order and effect of these events further.

We envision that the propensity of DnaA to open DNA could be adjusted in other bacterial species by strengthening or weakening the association of its ATPase domains with DNA and/or each other. An attractive feature of such a mechanism is that it is amenable to additional layers of control by changes to DUE sequence, superhelical density and co-resident architectural factors to ensure that a replication origin fires only when DnaA is both present and assembled properly. Such flexibility may have had a role in allowing DnaA to persist as the primary initiator in bacteria that have adapted to markedly different environmental niches.

Thematic patterns of substrate recognition in AAA+ ATPases

The mechanism by which DnaA coordinates ssDNA also comports well with findings in other replication initiation systems and with ASCE ATPases in general. For instance, many oligomeric RecA and

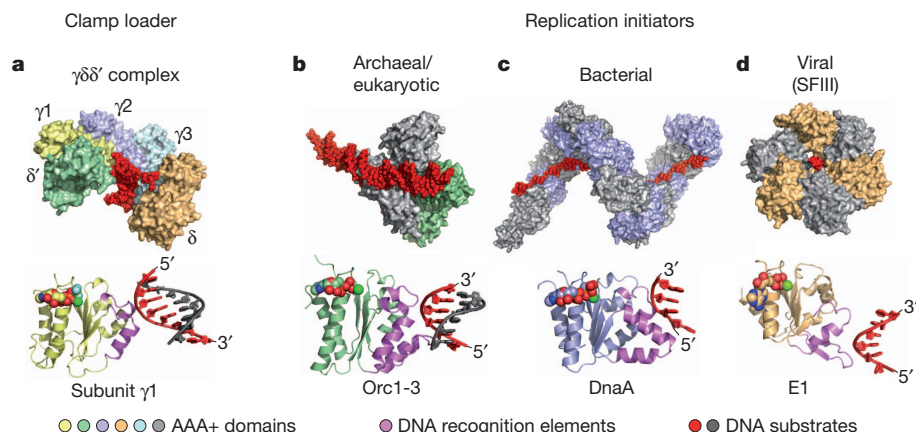


Figure 5 | Common DNA recognition strategies of AAA+ proteins. Structures of DNA-bound assemblies (top) and individual domains (bottom) for AAA+ proteins involved in replication. All recognize DNA using the same face of the AAA+ fold (violet; bottom). **a**, Bacterial clamp-loader ($\gamma\delta\delta'$) complex (AAA+ domains, differentially coloured) bound to primer-template DNA (PDB 3GLF)⁴¹. **b**, Archaeal initiators Orc1-1 (grey) and Orc1-3 (AAA+ domain, green) bound to origin DNA (PDB 2QBY)²⁶. **c**, Bacterial initiator DnaA (AAA+ domains, grey/blue) bound to ssDNA. **d**, Viral initiator/helicase E1 (AAA+ domains, orange/grey) bound to ssDNA (PDB 2GXA)⁴⁰. For all panels, DNA is shown as either red spheres (top) or as a red/grey cartoon (bottom). Nucleotide co-factors bound to AAA+ domains (bottom) are represented as spheres coloured by atom.

AAA+ enzymes bind substrate in the interior pore of a closed- or cracked-ring particle^{33,39–41}. DnaA follows this pattern. A comparison of DnaA to other, disparate nucleic-acid-dependent AAA+ systems—for example, polymerase clamp-loaders and processive helicases—further shows that these factors also associate with client substrates in a remarkably analogous manner, using the same face of the core $\alpha\beta\alpha$ ATP-binding fold to engage a short backbone stretch of their target DNAs (Fig. 5). For AAA+ proteins involved in initiation, these similar contact mechanisms have been differentially co-opted to assist with specific protein functions, ranging from the control of origin recognition (as seen in archaeal Orc1 proteins^{26,42}) to mediating processive DNA unwinding (viral superfamily 3 helicases^{40,43}). DnaA, with its ability to melt (but not translocate along) DNA, seems to use an intriguing mix of some of the activities exhibited by related initiation systems. Future efforts will be needed to determine how subtle differences in the position and nature of substrate-binding surfaces, combined with specific alterations in the assembly patterns of central AAA+ domains, endow such molecular motors and switches with their distinct biochemical properties.

METHODS SUMMARY

Detailed information regarding experimental methods, substrate sequences, binding constants, and FRET efficiencies and distances can be found in the Methods and in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 March; accepted 12 August 2011.

Published online 2 October 2011.

- Kornberg, A. & Baker, T. A. in *DNA Replication* (W. H. Freeman and Company, 1992).
- Stillman, B. Origin recognition and the chromosome cycle. *FEBS Lett.* **579**, 877–884 (2005).
- Duderstadt, K. E. & Berger, J. M. AAA+ ATPases in the initiation of DNA replication. *Crit. Rev. Biochem. Mol. Biol.* **43**, 163–187 (2008).
- Lee, D. G. & Bell, S. P. ATPase switches controlling DNA replication initiation. *Curr. Opin. Cell Biol.* **12**, 280–285 (2000).
- Leipe, D. D., Koonin, E. V. & Aravind, L. Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.* **333**, 781–815 (2003).
- Katayama, T., Ozaki, S., Keyamura, K. & Fujimitsu, K. Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and *oriC*. *Nature Rev. Microbiol.* **8**, 163–170 (2010).
- Kaguni, J. M. DnaA: controlling the initiation of bacterial DNA replication and more. *Annu. Rev. Microbiol.* **60**, 351–371 (2006).
- Leonard, A. C. & Grimwade, J. E. Regulating DnaA complex assembly: it is time to fill the gaps. *Curr. Opin. Microbiol.* **13**, 766–772 (2010).
- Fuller, R. S., Funnell, B. E. & Kornberg, A. The dnaA protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. *Cell* **38**, 889–900 (1984).
- Funnell, B. E., Baker, T. A. & Kornberg, A. *In vitro* assembly of a prepriming complex at the origin of the *Escherichia coli* chromosome. *J. Biol. Chem.* **262**, 10327–10334 (1987).
- Crooke, E., Thresher, R., Hwang, D. S., Griffith, J. & Kornberg, A. Replicatively active complexes of DnaA protein and the *Escherichia coli* chromosomal origin observed in the electron microscope. *J. Mol. Biol.* **233**, 16–24 (1993).
- Bramhill, D. & Kornberg, A. Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. *Cell* **52**, 743–755 (1988).
- Kowalski, D. & Eddy, M. J. The DNA unwinding element: a novel, cis-acting component that facilitates opening of the *Escherichia coli* replication origin. *EMBO J.* **8**, 4335–4344 (1989).
- Erzberger, J. P., Mott, M. L. & Berger, J. M. Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nature Struct. Mol. Biol.* **13**, 676–683 (2006).
- Ozaki, S. *et al.* A common mechanism for the ATP-DnaA-dependent formation of open complexes at the replication origin. *J. Biol. Chem.* **283**, 8351–8362 (2008).
- Speck, C. & Messer, W. Mechanism of origin unwinding: sequential binding of DnaA to double- and single-stranded DNA. *EMBO J.* **20**, 1469–1476 (2001).
- Sutton, M. D., Carr, K. M., Vicente, M. & Kaguni, J. M. *Escherichia coli* DnaA protein. The N-terminal domain and loading of DnaB helicase at the *E. coli* chromosomal origin. *J. Biol. Chem.* **273**, 34255–34262 (1998).
- Fang, L., Davey, M. J. & O'Donnell, M. Replisome assembly at *oriC*, the replication origin of *E. coli*, reveals an explanation for initiation sites outside an origin. *Mol. Cell* **4**, 541–553 (1999).
- Mott, M. L., Erzberger, J. P., Coons, M. M. & Berger, J. M. Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators. *Cell* **135**, 623–634 (2008).
- Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**, 27–43 (1999).
- Ogura, T. & Wilkinson, A. J. AAA+ superfamily ATPases: common structure—diverse function. *Genes Cells* **6**, 575–597 (2001).
- Jeruzalmi, D., O'Donnell, M. & Kuriyan, J. Crystal structure of the processivity clamp loader gamma (γ) complex of *E. coli* DNA polymerase III. *Cell* **106**, 429–441 (2001).
- Speck, C., Chen, Z., Li, H. & Stillman, B. ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. *Nature Struct. Mol. Biol.* **12**, 965–971 (2005).
- Clarey, M. G. *et al.* Nucleotide-dependent conformational changes in the DnaA-like core of the origin recognition complex. *Nature Struct. Mol. Biol.* **13**, 684–690 (2006).
- Duderstadt, K. E. *et al.* Origin remodeling and opening in bacteria rely on distinct assembly states of the DnaA initiator. *J. Biol. Chem.* **285**, 28229–28239 (2010).
- Dueber, E. L., Corn, J. E., Bell, S. D. & Berger, J. M. Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. *Science* **317**, 1210–1213 (2007).
- Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **146**, 11–31 (2004).
- Story, R. M., Weber, I. T. & Steitz, T. A. The structure of the *E. coli* recA protein monomer and polymer. *Nature* **355**, 318–325 (1992).
- Kowalczykowski, S. C. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* **25**, 156–165 (2000).
- Cox, M. M. Motoring along with the bacterial RecA protein. *Nature Rev. Mol. Cell Biol.* **8**, 127–138 (2007).
- Egelman, E. A common structural core in proteins active in DNA recombination and replication. *Trends Biochem. Sci.* **25**, 179–182 (2000).
- Conway, A. B. *et al.* Crystal structure of a Rad51 filament. *Nature Struct. Mol. Biol.* **11**, 791–796 (2004).
- Chen, Z., Yang, H. & Pavletich, N. P. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* **453**, 489–494 (2008).
- Stasiak, A. & Di Capua, E. The helicity of DNA in complexes with recA protein. *Nature* **299**, 185–186 (1982).
- Galletto, R., Amitani, I., Baskin, R. J. & Kowalczykowski, S. C. Direct observation of individual RecA filaments assembling on single DNA molecules. *Nature* **443**, 875–878 (2006).
- Nishinaka, T., Ito, Y., Yokoyama, S. & Shibata, T. An extended DNA structure through deoxyribose-base stacking induced by RecA protein. *Proc. Natl Acad. Sci. USA* **94**, 6623–6628 (1997).
- Joo, C. *et al.* Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* **126**, 515–527 (2006).
- Bianchi, M., Riboli, B. & Magni, G. *E. coli* recA protein possesses a strand separating activity on short duplex DNAs. *EMBO J.* **4**, 3025–3030 (1985).
- Thomsen, N. D. & Berger, J. M. Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* **139**, 523–534 (2009).
- Enemark, E. J. & Joshua-Tor, L. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**, 270–275 (2006).
- Simonetta, K. R. *et al.* The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. *Cell* **137**, 659–671 (2009).
- Gaudier, M., Schuwirth, B. S., Westcott, S. L. & Wigley, D. B. Structural basis of DNA replication origin recognition by an ORC protein. *Science* **317**, 1213–1216 (2007).
- Liu, X., Schuck, S. & Stenlund, A. Adjacent residues in the E1 initiator β -hairpin define different roles of the beta-hairpin in Ori melting, helicase loading, and helicase activity. *Mol. Cell* **25**, 825–837 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank K. Drlica, J. Keck, T. Murray and the Berger laboratory for helpful comments, and M. M. Cox for his contribution of RecA protein. This work was supported by the NIGMS (GM071747) and the National Institute of Health Molecular Biophysics Training Grant T32 GM008295.

Author Contributions K.E.D. and J.M.B. designed the experiments, analysed the data and wrote the paper. Protein purification, crystallization and ssDNA binding assays were performed by K.C. and K.E.D. K.E.D. performed the other experiments.

Author Information Coordinates have been deposited in the RSCB Protein Data Bank under the accession number 3R8F. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.M.B. (jmberger@berkeley.edu).

METHODS

Expression and purification of DnaA. Residues 76–399 of *A. aeolicus* DnaA (containing the AAA+ and duplex-DNA-binding regions) were expressed as a TEV-protease cleavable His₆-MBP fusion and purified as previously described²⁵. As a final purification step, untagged DnaA proteins (from TEV cleavage) were run over an S-200 size-exclusion column (GE) in gel-filtration buffer (50 mM HEPES pH 7.5, 500 mM KCl, 10% (v/v) glycerol, 5 mM MgCl₂, 100 μM ADP). Monomeric species were pooled, concentrated and flash-frozen for storage at –80 °C. For mutagenesis studies, changes were introduced into the His₆-MBP–DnaA construct using QuickChange (Stratagene).

Crystallization and DNA soaking. After gel filtration of DnaA in crystallization buffer (20 mM HEPES pH 7.5, 250 mM KCl, 250 mM KBr, 10% (v/v) glycerol, 10 mM MgCl₂, 100 μM AMPPCP), monomeric species were pooled, concentrated to 10 mg ml^{–1} at 4 °C, and flash-frozen for storage at –80 °C. Crystallization by hanging-drop vapour diffusion was performed by mixing 1.3 μl of freshly thawed DnaA in crystallization buffer and 1 μl of well solution (15–35 mM sodium cacodylate pH 6.5, 26% 1,2-propanediol and 1–2% PEG 2000 MME) at 18 °C. Large rod-like crystals appeared within 1 to 2 weeks and reached maximal size around 3 weeks. Crystals were transferred by looping to a low-salt soaking solution (20 mM HEPES pH 7.5, 30 mM sodium cacodylate pH 6.5, 10% (v/v) glycerol, 10 mM MgCl₂, 26% 1,2-propanediol, 2.5% PEG 2000 MME and 200 μM AMPPCP) containing 5 mM ssDNA. After 6 h, crystals were looped and transferred to a second drop of soaking solution containing 5 mM ssDNA, and left overnight to ensure both complete removal of remaining salt and to allow time for binding. The crystals were then looped and flash frozen in liquid nitrogen in preparation for data collection. Previous biochemical studies revealed no apparent sequence preference for ssDNA by *A. aeolicus* DnaA²⁵, so DNAs of various sequences and lengths were all individually tested (dT_n (*n* = 3–12) and dA_n (*n* = 3–12), Elim Biopharmaceuticals). Data collection and structure determination revealed that dA₁₂ generated the strongest electron density, although similar, albeit weaker and less connected, density was observed for dT oligonucleotides and smaller dA substrates.

Data collection and structure determination. Data were collected at Beamline 8.3.1 at the Advanced Light Source (ALS)⁴⁴ and processed using HKL-2000⁴⁵. Crystals belong to the space group P2₁2₁2₁, with dA₁₂ soaked crystals having unit cell dimensions *a* = 99.8 Å, *b* = 114.2 Å and *c* = 201.3 Å (Supplementary Table 1). Data were phased using molecular replacement as implemented in PHENIX⁴⁶, using a DNA-free DnaA tetramer as a search model (PDB 2HCB)¹⁴. Initial *F*_o – *F*_c electron density maps containing clear density for DNA were generated using rigid-body and grouped B-factor refinement with PHENIX⁴⁶. Further refinement was conducted using multibody, non-crystallographic-symmetry (NCS)-restrained, simulated annealing in PHENIX⁴⁶, fourfold multibody NCS averaging with a custom solvent mask (including the region of DNA binding), density modification using resolve⁴⁷, and manual model building in COOT⁴⁸. During the final stages of refinement, fourfold multibody NCS and secondary structure restraints were retained for AMPPCP and the entire protein except residues 255–265, which differed between chains as a result of crystal packing interactions. Composite, simulated-annealing omit maps generated with CNS⁴⁹ were used as a guide for building with COOT. DNA and waters were manually added to the model, and final rounds of refinement with PHENIX were conducted with grouped B-factor modelling, as well as NCS restraints and TLS modelling of individual protein domains (comprising three TLS groups in total: the AAA+-core (amino acids 76–241) plus AMPPCP; the AAA+ α-helical 'lid' (amino acids 242–254 and 266–241); and the duplex-DNA binding domain (amino acids 291–399)). All panels of figures with renderings of structures and electron density were prepared with PyMol⁵⁰.

The final model contains one DnaA tetramer bound to one dA₁₂ per asymmetric unit. A clear 5' or 3' break between successive dA₁₂ substrates was not present in the electron density, indicating that during the soaking procedure, different DNA molecules bound in multiple registers to consecutive DnaA protomers throughout the crystal. Accordingly, a terminal 5' phosphate, which was not present in the substrate used for soaking, was added to the modelled dA₁₂ DNA.

Polarity of DNA binding. During refinement, DNA was initially modelled independently into the DnaA pore with each of the two possible polarities. Compared to the 5' to 3' polarity presented in the paper, refinement of the model with the dA₁₂ substrate running 3' to 5' (from the arginine finger side to the nucleotide-binding face of a DnaA protomer) resulted in only marginally higher *R*_{work} and *R*_{free} values (~0.1%), but also the appearance of off-model positive difference density and on-model negative difference density in *F*_o – *F*_c maps (the model as presented in the paper did not display such features). Simultaneous refinement with two DNAs, each at half occupancy, with opposing polarities of the dA₁₂

substrate resulted in ~0.3% higher *R*_{work} and *R*_{free} values, and again showed unfavourable difference density in *F*_o – *F*_c maps.

Recognizing that these differences, although consistent with our build, were subtle and did not definitively resolve the ssDNA binding polarity to DnaA, we set out to test our assignment further. To this end, we designed and had synthesized (by Trilink BioTechnologies) two specialized, di-adenosyl nucleotide substrates that would give rise to a clear distinction in binding orientation: 5'-p(Br-A)pAp and 5'-p(ε-A)pAp, where 'Br-A' indicates a bromo-deoxyadenosine label, 'ε-A' indicates an etheno-deoxyadenosine label, and 'p' indicates a phosphate moiety. Soaking of crystals with these dinucleotide substrates was performed as described for ssDNA substrates (note that in our soaking trials with oligonucleotides as short as dA₃, we observed density associated with DnaA protomers consistent with that seen for the trinucleotide repeats when using dA₁₂). Unfortunately, data collected with the ε-A-substituted oligonucleotide yielded maps with density for dinucleotides bound to each monomer but additional density for the EthenoA was not clearly visible, probably due to the low ~3.4 Å resolution limit of the DnaA crystals. At the same time, SAD data sets collected with the Br-A-substituted oligonucleotide did not yield useful maps, due to the weak diffraction of the crystals (and accompanying radiation damage as we attempted to maximize data signal-to-noise at the bromine absorption maximum⁵¹), to incomplete bromine labelling, or both. We note that we carried out soaks with longer Br-dA-labelled (and Br-dU-labelled) oligonucleotides, but these efforts were not successful, again because of weak diffraction. Additional experiments to test the orientation (for example, using labelled oligo/protein pairs and FRET) were considered, but ruled out due to the small binding site size for substrate DNA, and an inability to find a suitable pair of labelling sites that could report on differing binding orientations.

As a consequence, although our data are supportive of the polarity presented in our model, we cannot definitively rule out the possibility that ssDNA might also be binding to DnaA in the crystal in an opposing direction. Nonetheless, several findings support the idea that DnaA binds ssDNA in a defined orientation that is consistent with the direction suggested here. For example, following nucleoprotein complex formation on *oriC*, DnaA melts (A+T)-rich regions in the DUE¹²; two independent reports have found that *E. coli* DnaA binds specifically to only one strand (the so-called 'top' strand) of the DUE during this process^{15,16}. The importance of DnaA binding polarity becomes clear during the next stage of initiation, when the DnaB helicase is loaded. Modelling studies based on the known DnaB translocation polarity (5' to 3') and known pairwise interactions between DnaB, DnaC and DnaA, have suggested that top-strand loading involves a direct interaction between DnaA and DnaC that has been observed biochemically and depends on the AAA+ domains of the two proteins¹⁹. Because AAA+ domains assemble with a defined orientation, in which the arginine finger face of one protomer points into the nucleotide binding face of a second subunit, it follows that DnaA molecules probably position themselves on the top strand with only one of their two AAA+ domain surfaces presented to DnaC. Although the polarity of the DnaA–DnaC interaction has not been established, a mutation on the arginine finger face of the *E. coli* DnaA AAA+ domain, R281A, is reported to disrupt helicase loading, but not *oriC* melting⁵²; this finding indicates that DnaA interacts with DnaC using its arginine-finger face. In our structures, the 5' end of the modelled DNA resides near the arginine finger face of DnaA, a configuration consistent with these data.

ssDNA extension assay. Extension of dT₂₁ oligonucleotides labelled with Cy3 and Cy5 (FR-dT₂₁) by DnaA was monitored by FRET using a FluoroMax-4 (Horiba Jobin Yvon) spectrofluorimeter. Measurements were carried out at 25 °C in 20 μl with 25 nM of FR-dT₂₁ and either 10 μM of DnaA in DnaA extension buffer (50 mM HEPES pH 7.5, 125 mM KCl, 2% (v/v) glycerol, 10 mM MgCl₂ and 2 mM ADP or ADP•BeF₃) or 10 μM of RecA in RecA extension buffer (25 mM Tris-acetate pH 7.5, 100 mM Na-acetate, 10 mM Mg-acetate, 1 mM DTT and 2 mM ADP or ATPγS). Emission scans from 545 to 700 nm were collected with excitation of Cy3 at 530 nm, divided by the excitation intensity, and then corrected for the wavelength-dependent sensitivity of the detector. FRET efficiencies and distances were determined by comparing the Cy3 fluorescence from the doubly labelled substrate (FR-dT₂₁) with the Cy3 fluorescence from a substrate only having a Cy3 label (C3-dT₂₁) under the same conditions.

Influence of proteins on dye behaviour. To ensure that all influences on dye behaviour were properly considered when processing the FRET data from the DNA extension assay (Fig. 3), the fluorescence and absorbance of each dye was monitored independently for each experimental condition. Emission and absorbance scans of substrates labelled only with the Cy3 donor (C3-dT₂₁) were collected in buffer alone, and with protein in the presence of ATP mimics or ADP (Supplementary Fig. 8c (panels ii and iii) and d (panels ii and iii)). Emission scans revealed pronounced protein- and nucleotide-dependent enhancement of donor fluorescence, but negligible differences in donor absorbance. Similar, but less significant, effects have been observed previously for RecA at a concentration

of 1 μM (comparable to our working concentration of 10 μM)³⁷. A similar enhancement in acceptor fluorescence (and lack of effect on acceptor absorbance) also was observed in the doubly labelled substrate, FR-dT₂₁ (Supplementary Fig. 8c (panels iv and v) and d (panels iv and v)). These effects are not surprising, as the spectral properties of fluorescent dyes are known to undergo marked variation depending on chemical environment^{53,54}. However, these controls also indicated that we needed to take into account additional corrections to obtain accurate distance measurements. In particular, the changes in donor fluorescence, but not donor absorbance, were indicative of changes in the donor quantum yield (Φ_D), which is used to calculate R_0 (\AA), the distance corresponding to a FRET efficiency of 50%: $R_0 = 8.79 \times 10^{-5} (J\kappa^2 n^{-4} \Phi_D)^{1/6}$ where J is the spectral overlap between the donor emission and acceptor absorption; κ^2 is a geometric factor that depends on the orientation of donor and acceptor; and n is the refractive index of the medium between donor and acceptor⁵⁴.

To determine the donor quantum yield under different experimental conditions, we used rhodamine 6G as a standard for calibration, with an assumed quantum yield of 0.95 in EtOH⁵⁵. To calculate the quantum yields seen in Supplementary Table 5, we collected the fluorescence and absorbance of the donor-only labelled substrate (C3-dT₂₁) in the RecA and DnaA buffers. We then used these data to determine ratios between the integrated fluorescence and absorbance, while correcting for the fractional absorbance at the excitation wavelengths used. Fluorescein in 0.1 M NaOH (known to have a quantum yield of 0.95 (ref. 56)) was also measured as a control. To ensure reliable readings, all absorbance measurements were conducted with 1 μM of the donor-only labelled substrate, either alone or in the presence of 10 μM of the indicated protein. All emission measurements were conducted with 25 nM of the donor-only labelled substrate, either alone or in the presence of 10 μM of the indicated protein. Because the presence of protein had no influence on dye absorbance (Supplementary Fig. 8c (panels iii and v)), the quantum yield of the donor in the presence of different proteins was determined simply by using its value in buffer, and multiplying by the observed changes in fluorescence. R_0 values were then calculated for each sample using the corresponding values for quantum yield.

Determination of FRET efficiencies and DNA length. To determine the efficiency of transfer (E) from the FRET data collected using the DNA extension assay (Fig. 3 and Supplementary Figs 6, 8 and 9), the emission of the donor from the donor-only labelled substrate (C3-dT₂₁, F_D) was compared to the emission of the donor from the doubly labelled substrate FR-dT₂₁ (F_{DA}) under equivalent experimental conditions as follows: $E = 1 - \frac{F_{DA}}{F_D}$ (ref. 57). The efficiencies for different samples can be found in Supplementary Table 5. Solution distances were subsequently obtained using the relation $R = R_0 \left[\frac{1-E}{E} \right]^{1/6}$ (ref. 54), the values of which can also be found in Supplementary Table 5.

DNA strand displacement assay. The DnaA-dependent displacement of single strands from duplex DNA was monitored using a Cy3 label on one of two strands (the 'bottom' strand, Supplementary Table 3). All measurements were carried out at 25 °C in 80 μl of binding buffer containing 50 mM HEPES pH 7.5, 125 mM KCl, 2% (v/v) glycerol, 10 mM MgCl₂, 0.1 mg ml⁻¹ bovine serum albumin, 1 mM DTT and 2 mM ADP or ADP•BeF₃ (a non-hydrolysable ATP analogue that mimics the properties of ATP^{25,58}). After a short 2-min incubation of 25 nM duplex DNA with various DnaA concentrations (Fig. 4), 50 nM of unlabelled bottom strand was added for an additional 30 min to capture displaced top strands. After quenching with 10 \times stop buffer containing 200 mM EDTA, 10 mg ml⁻¹ proteinase K and 4% (v/w) SDS, displaced strands were separated on native polyacrylamide gels in Tris/boric acid/EDTA (TBE) buffer and visualized using a Molecular Dynamics Typhoon. The time dependence of DNA strand displacement by DnaA can be found in Supplementary Fig. 11. Sequences of substrates used can be found in Supplementary Table 3.

ssDNA binding assay. Binding of 5' fluorescein-labelled dT₂₅ oligonucleotides (F-dT₂₅, Supplementary Table 3) to DnaA was monitored by fluorescence polarization

using a Victor 3V (Perkin Elmer) multi-label plate reader (Supplementary Fig. 4a). Measurements were carried out at 25 °C in 20 μl of binding buffer containing 50 mM HEPES pH 7.5, 125 mM KCl, 2% (v/v) glycerol, 10 mM MgCl₂, 0.1 mg ml⁻¹ bovine serum albumin, 1 mM DTT and 2 mM ADP•BeF₃. The concentration of F-dT₂₅ was held constant at 10 nM while the concentration of DnaA was varied. All data points represent the average of three independent measurements, with error bars representing the standard deviation between measurements. Binding curves were fit to the Hill equation to obtain $K_{d,app}$ values (Supplementary Table 4) as described previously²⁵.

Oligomerization characteristics of ssDNA-binding mutants. To confirm that ssDNA binding mutations did not affect the ATP-dependent oligomerization properties of DnaA, we used a previously established glutaraldehyde-crosslinking assay²⁵. Crosslinking was performed by incubating 50 μg ml⁻¹ of various *A. aeolicus* DnaA proteins in 80 μl of a reaction buffer (50 mM HEPES pH 7.5, 10% (v/v) glycerol, 125 mM KCl, 5 mM MgCl₂, 2 mM DTT) containing 2 mM ADP•BeF₃ at 25 °C for 5 min. Glutaraldehyde (Polysciences Inc.) was then added to 1 mM final concentration using 8.8 μl of a 10 mM stock. Reactions were incubated at 25 °C for an additional 1 min before quenching with 8 μl of 200 mM glycine followed by the addition of 30 μl of gel loading buffer (100 mM Tris pH 6.8, 24% (v/v) glycerol, 8% (w/v) SDS, 200 mM DTT, 0.02% (w/v) bromophenol blue). Crosslinked proteins were loaded in a volume of 15 μl and separated on denaturing 4.5% polyacrylamide gels (80:1 acrylamide:bisacrylamide) in 0.1 M sodium phosphate, 0.1% SDS buffer (pH 7.2)^{59,60}, and visualized by silver staining (Supplementary Fig. 4b).

44. MacDowell, A. A. *et al.* Suite of three protein crystallography beamlines with single superconducting bend magnet as the source. *J. Synchrotron Radiat.* **11**, 447–455 (2004).
45. Otwinowski, Z. & Minor, W. in *Methods in Enzymology* 307–326 (Academic Press, 1997).
46. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
47. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
49. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
50. DeLano, W. L. The PyMOL Molecular Graphics System (DeLano Scientific, 2002).
51. Ennifar, E., Carpentier, P., Ferrer, J. L., Walter, P. & Dumas, P. X-ray-induced debromination of nucleic acids at the Br K absorption edge and implications for MAD phasing. *Acta Crystallogr. D* **58**, 1262–1268 (2002).
52. Felczak, M. M. & Kaguni, J. M. The box VII motif of *Escherichia coli* DnaA protein is required for DnaA oligomerization at the *E. coli* replication origin. *J. Biol. Chem.* **279**, 51156–51162 (2004).
53. Mujumdar, R. B., Ernst, L. A., Mujumdar, S. R., Lewis, C. J. & Waggoner, A. S. Cyanine dye labeling reagents: sulfoindocyanine succinimidyl esters. *Bioconjug. Chem.* **4**, 105–111 (1993).
54. Cantor, C. R. & Schimmel, P. R. in *Biophysical Chemistry: Part II: Techniques for the study of biological structure and function* 846 (W. H. Freeman and Company, 1980).
55. Magde, D., Wong, R. & Seybold, P. G. Fluorescence quantum yields and their relation to lifetimes of rhodamine 6G and fluorescein in nine solvents: improved absolute standards for quantum yields. *Photochem. Photobiol.* **75**, 327–334 (2002).
56. Lakowicz, J. R. in *Principles of Fluorescence Spectroscopy* 2nd edn. (Kluwer/Plenum, 1999).
57. Clegg, R. M. Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol.* **211**, 353–388 (1992).
58. Petsko, G. A. Chemistry and biology. *Proc. Natl Acad. Sci. USA* **97**, 538–540 (2000).
59. Crisona, N. J. & Cozzarelli, N. R. Alteration of *Escherichia coli* topoisomerase IV conformation upon enzyme binding to positively supercoiled DNA. *J. Biol. Chem.* **281**, 18927–18932 (2006).
60. Weber, K. & Osborn, M. The reliability of molecular weight determinations by dodecyl sulfate-polyacrylamide gel electrophoresis. *J. Biol. Chem.* **244**, 4406–4412 (1969).

Active tactile exploration using a brain–machine–brain interface

Joseph E. O'Doherty^{1,2}, Mikhail A. Lebedev^{2,3}, Peter J. Ifft^{1,2}, Katie Z. Zhuang^{1,2}, Solaiman Shokur⁴, Hannes Bleuler⁴ & Miguel A. L. Nicolelis^{1,2,3,5,6}

Brain–machine interfaces^{1,2} use neuronal activity recorded from the brain to establish direct communication with external actuators, such as prosthetic arms. It is hoped that brain–machine interfaces can be used to restore the normal sensorimotor functions of the limbs, but so far they have lacked tactile sensation. Here we report the operation of a brain–machine–brain interface (BMBI) that both controls the exploratory reaching movements of an actuator and allows signalling of artificial tactile feedback through intracortical microstimulation (ICMS) of the primary somatosensory cortex. Monkeys performed an active exploration task in which an actuator (a computer cursor or a virtual-reality arm) was moved using a BMBI that derived motor commands from neuronal ensemble activity recorded in the primary motor cortex. ICMS feedback occurred whenever the actuator touched virtual objects. Temporal patterns of ICMS encoded the artificial tactile properties of each object. Neuronal recordings and ICMS epochs were temporally multiplexed to avoid interference. Two monkeys operated this BMBI to search for and distinguish one of three visually identical objects, using the virtual-reality arm to identify the unique artificial texture associated with each. These results suggest that clinical motor neuroprostheses might benefit from the addition of ICMS feedback to generate artificial somatic perceptions associated with mechanical, robotic or even virtual prostheses.

Brain–machine interfaces (BMIs) have evolved from 1-d.f. systems³ to many-d.f. robotic arms⁴ and muscle stimulators⁵ that perform complex limb movements, such as reaching^{6–8} and grasping⁹. However, somatosensory feedback, which is essential for dexterous control^{10–12}, remains underdeveloped in BMIs. With the exception of a few studies combining BMIs with tactile stimuli applied to the body¹³, existing systems rely almost exclusively on visual feedback. Prosthetic sensation has been studied in the context of sensory substitution¹⁴ and targeted reinnervation¹⁵; however, these approaches have limited application range and channel capacity. To provide a proof-of-concept method of equipping neuroprostheses with sensory capabilities, we implemented a BMBI that extracts movement commands from the motor areas of the brain while delivering ICMS feedback in somatosensory areas^{1,2,16} to evoke discriminable percepts^{17–20}. This idea received support from our pilot study¹⁶, in which a monkey responded to ICMS cues with the movements of a BMI-controlled cursor. However, the ICMS cue did not provide feedback of object–actuator interactions in this previous demonstration.

The BMBI developed here allowed active tactile exploration²¹ during BMI control (Fig. 1a). Two monkeys (M and N) received multielectrode implants in the primary motor cortex (M1) and the primary somatosensory cortex (S1) (Fig. 1b). They explored virtual objects using either a computer cursor or a virtual image (avatar) of an arm (Supplementary Fig. 1a, b). In ‘hand control’, the monkeys moved a joystick with their left hands to position the actuator. They searched through a set of virtual objects, selected one with a particular artificial

texture conveyed by ICMS, and held the actuator over that object to obtain reward (Fig. 1a and Supplementary Fig. 1c, d). During ‘brain control’, the joystick was disconnected and the actuator was controlled by the activity of right-hemisphere M1 neurons^{9,22,23}. The behavioural tasks varied in the number of objects on the screen, the artificial textures used and the actuator type (Fig. 2a), and were more difficult than previously reported BMI tasks because of the presence of multiple objects in the workspace, a prolonged object selection period and the necessity of interpreting ICMS feedback.

ICMS was delivered through two pairs of microwires to the hand representation area of S1 in monkey M (Fig. 1c) and through one pair

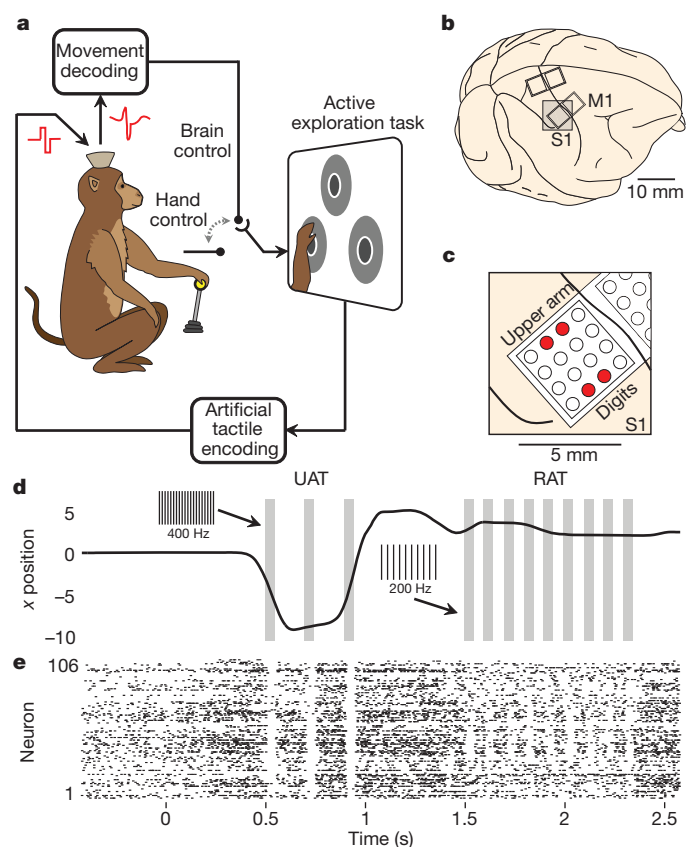


Figure 1 | The brain–machine–brain interface. **a**, Movement intentions are decoded from M1; artificial tactile feedback is delivered to S1. **b**, Microwires were implanted in M1 and S1. **c**, Microwires used for ICMS in monkey M are accentuated in red. **d**, Actuator movements for a trial in which monkey M explores UAT but ultimately selects RAT. Grey bars indicate stimulation patterns; insets indicate the ICMS frequency. **e**, Rastergram of M1 neurons recorded during the trial shown in **d**.

¹Department of Biomedical Engineering, Duke University, Durham, North Carolina 27708, USA. ²Center for Neuroengineering, Duke University, Durham, North Carolina 27710, USA. ³Department of Neurobiology, Duke University, Durham, North Carolina 27710, USA. ⁴STI IMT, École Polytechnique Fédérale de Lausanne, Lausanne CH1015, Switzerland. ⁵Department of Psychology and Neuroscience, Duke University, Durham, North Carolina 27708, USA. ⁶Edmond and Lily Safra International Institute of Neuroscience of Natal, Natal 59066-060, Brazil.

of microwires to the leg representation in monkey N. Each artificial texture consisted of a high-frequency pulse train presented in packets at a lower, secondary, frequency (Fig. 1d and Supplementary Fig. 2a). The rewarded artificial texture (RAT) consisted of 200-Hz pulse trains delivered in 10-Hz packets. The comparison artificial textures comprised 400-Hz pulse trains delivered in 5-Hz packets (unrewarded artificial texture (UAT)) or an absence of ICMS (null artificial texture (NAT)).

The main challenge solved here was the real-time coupling of ICMS feedback to the BMI decoder. Because ICMS artefacts masked neuronal activity for 5–10 ms after each pulse (Fig. 1d, e), we multiplexed neuronal recordings and ICMS with a 20-Hz clock rate (Supplementary Fig. 2a). The interleaved intervals proved adequate for online motor control and artificial sensation—a result that was not clear a priori because S1 stimulation could have affected M1 processing through the connections between these areas.

BMBI performance improved with training. In task I (Fig. 2a), monkey M surpassed chance performance after nine sessions and monkey N did so after four sessions ($P < 0.001$, one-sided binomial test). Improvement continued with more difficult tasks (tasks II–V) (Fig. 2a, b and Supplementary Fig. 3a). In particular, the time spent exploring unrewarded artificial textures decreased (Fig. 2c and Supplementary Fig. 3b). Additionally, performance improved over the course of daily experimental sessions (Fig. 2d). Psychometric analysis of RAT stimulation amplitudes (Supplementary Fig. 2b) indicated that at least 8 nC per ICMS waveform phase (100- μ s-wide current pulses of

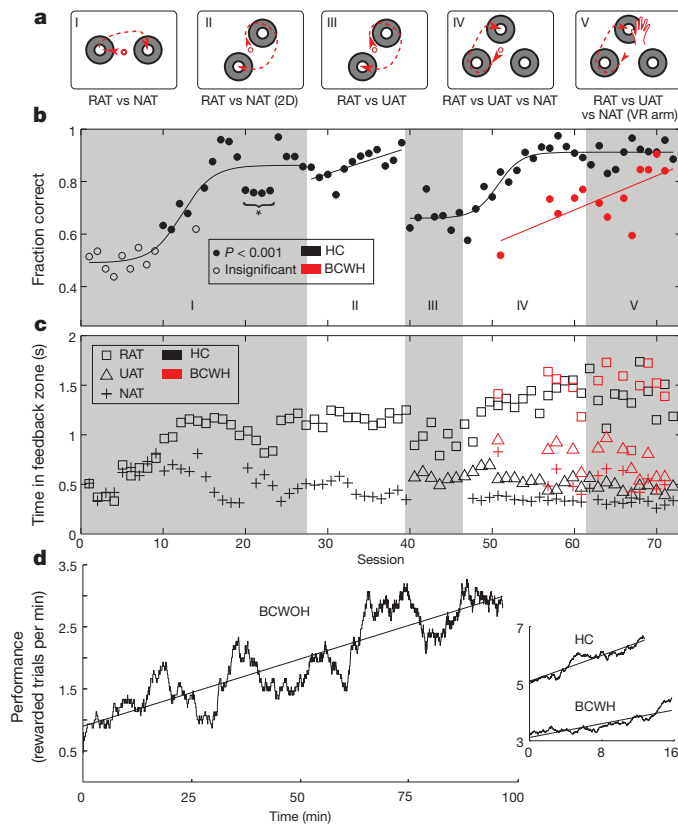


Figure 2 | Learning to use ICMS feedback. **a**, Behavioural tasks. 2D, two-dimensional; VR, virtual reality. **b**, **c**, Performance of monkey M (72 sessions). Circles (**b**) depict the fraction of correctly performed trials. Open circles indicate chance performance. Curves are lines of best fit. The asterisk indicates sessions used for psychometric measurements. Squares, triangles and crosses (**c**) represent mean times spent in RAT, UAT and NAT, respectively. Black, hand control; red, brain control. **d**, Intrasession performance for monkey M. Curves represent averages for brain control without hand movements (BCWOH; main panel, three sessions), for hand control (inset, 12 sessions) and for brain control with hand movements (BCWH; inset, 12 sessions). Lines are best linear fits.

80 μ A) was needed for the discrimination of artificial textures ($P < 0.001$, one-sided binomial test). Performance was at chance level for catch trials (task II), where ICMS was not delivered ($P = 0.90$, one-sided binomial test).

The statistics of object exploration intervals (total time spent over a particular object in a given trial) indicated that the monkeys uniquely discriminated each type of artificial texture (Figs 2c and 3a, c) and interpreted ICMS within hundreds of milliseconds—a timescale comparable to that for the discrimination of peripheral tactile stimuli^{24,25}. Early in task I, exploration intervals were equal for RAT and NAT ($P > 0.5$, Wilcoxon signed-rank test); with training, they became longer for RAT and shorter for NAT (tasks I and II) and UAT (tasks III–V). During hand control, the mean interval was longest for RAT (monkey M: $1,396 \pm 21$ ms; monkey N: $1,165 \pm 15$ ms; mean \pm s.e.m.), shortest for NAT (304 ± 8 ms; 300 ± 10 ms) and intermediate for UAT (452 ± 13 ms; 402 ± 14 ms) ($P < 0.01$, analysis of variance). During brain control, intervals spent exploring NAT (498 ± 15 ms; 587 ± 25 ms) and UAT (685 ± 20 ms; 764 ± 32 ms) were longer than they were during hand control, but were still shorter than those spent exploring RAT ($1,420 \pm 28$ ms; $1,398 \pm 55$ ms) ($P < 0.01$, analysis of variance).

Additional hallmarks of active exploration were seen in the conditional probabilities of selecting different artificial textures (Fig. 3b, d). During hand control trials, the monkeys stayed over the first-encountered artificial texture (arrows that loop back to the same

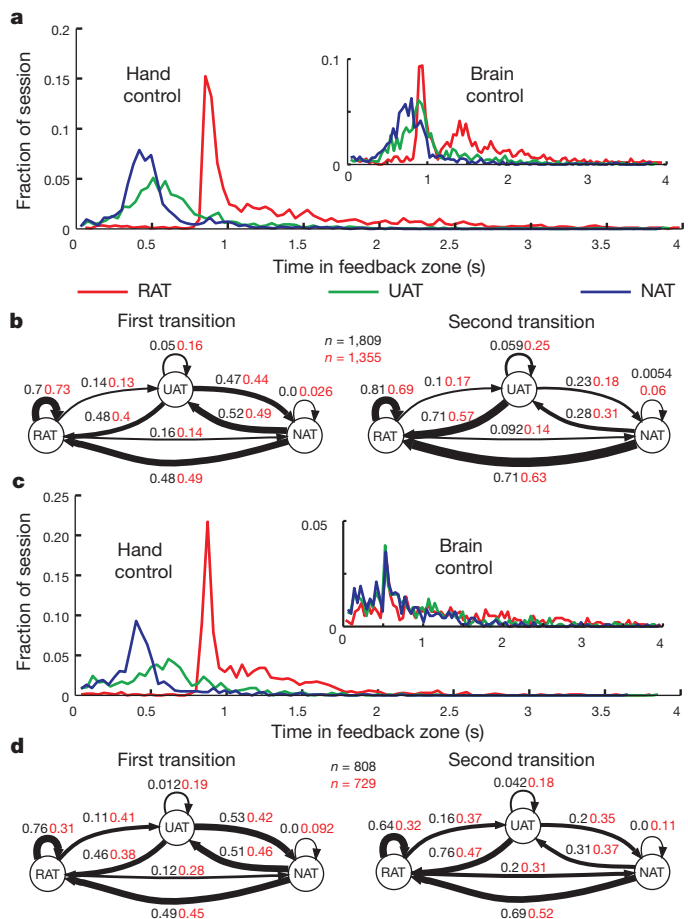


Figure 3 | Statistics of object exploration. **a**, Object exploration intervals during hand control and brain control (inset) for monkey M (hand control: $n = 1,809$ trials; brain control: $n = 1,355$ trials). **b**, State transition diagrams for monkey M, indicating the probabilities of reaching among RAT, UAT and NAT after the first (left subpanel) or second (right subpanel) reach. Black labels, hand control; red labels, brain control; line thickness is proportional to transition probability. **c**, Same as **a**, but for monkey N (hand control: $n = 808$ trials; brain control: $n = 729$ trials). **d**, Same as **b**, but for monkey N.

artificial texture in Fig. 3b, d) with high probability if it was RAT (monkey M: $P = 0.70$; monkey N: $P = 0.76$), but with low probability if it was UAT ($P = 0.05$; $P = 0.01$) or NAT ($P = 0.0$; $P = 0.0$) (Fig. 3b, d, left). After examining the second artificial texture, the monkeys could identify the correct artificial texture either by apprehending it directly or through a process of elimination. This follows from the increase from chance to approximately $P = 0.7$ in the probability of moving to RAT from NAT or UAT and the decrease to $P \approx 0.2$ in the probability of revisiting UAT or NAT (Fig. 3b, d, right). Similar effects were observed for brain control (Fig. 3b, d, red text).

Brain control started in task IV. During BCWH, the monkeys continued to hold the joystick although it was disconnected^{16,22}. During BCWOH^{9,22}, the joystick was removed. In monkey M, with more than 200 recorded neurons, performance was less accurate during BCWH ($73.75 \pm 3.00\%$; mean \pm s.e.m.) than during hand control ($91.48 \pm 1.20\%$). In monkey N, with 50 recorded neurons, performance dropped further ($50.37 \pm 3.74\%$ versus $91.45 \pm 1.91\%$), but still significantly exceeded the 33% chance level. M1 neurons showed directionally tuned modulations (Supplementary Figs 5 and 6) that were retained across different interfering ICMS patterns during both hand control (Supplementary Fig. 4a, b) and brain control (Fig. 4a, b).

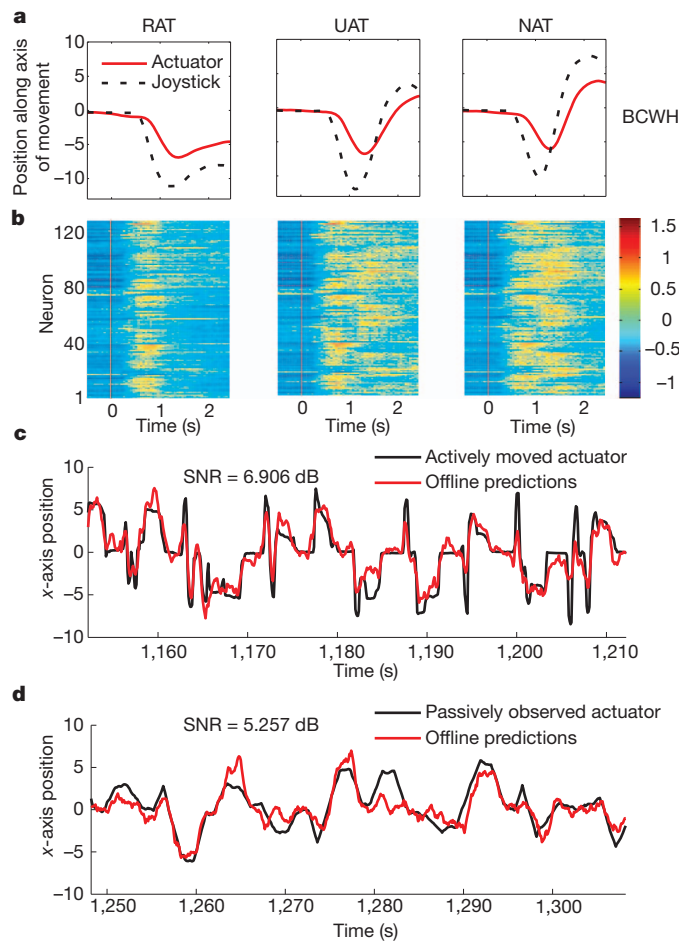


Figure 4 | M1 modulations during active control versus passive observation. **a, b,** Average brain control movements for monkey M ($n = 294$ trials) towards the RAT, UAT and NAT objects appearing on the left-hand side of the screen (**a**), and corresponding neuronal modulations (**b**). The colour scale shows normalized firing rate (Hz). Only trials with subsequent anticlockwise reaching movements are included in the middle and right subpanels. Red vertical lines indicate object onset. Firing rate normalized by s.d. **c, d,** Position of the actuator actively controlled (**c**) and passively observed (**d**) by monkey M. SNR, signal-to-noise ratio.

In BCWOH, task requirements were eased: the object selection period was reduced to 300–500 ms and monkeys were allowed to overstay at an incorrect object. The performance of monkey M, measured as the number of rewards per minute, steadily improved from 1.021 ± 0.007 to 2.962 ± 0.005 (mean \pm s.e.m.; Fig. 2d). Similar improvements were observed for hand control and BCWH (Fig. 2d, inset). The average frequency of actuator displacements, calculated from power spectra, was correlated with the improvement in performance during BCWOH ($R^2 = 0.16$ for the horizontal (x) coordinate and $R^2 = 0.26$ for the vertical (y) coordinate; $P < 0.001$, F -test), which indicated that the monkey modulated its brain activity to scan the targets faster. This behaviour was not random, as the exploration interval for NAT ($3,620 \pm 350$ ms; mean \pm s.e.m.) was significantly shorter ($P < 0.02$, Wilcoxon rank-sum test) than for UAT ($4,270 \pm 310$ ms). The exploration of RAT ($2,255 \pm 94$ ms) was the shortest owing to the reduced selection period. For monkey N, BCWOH performance (2.084 ± 0.085 rewards per minute) did not change within sessions, and the differences in exploration intervals were not significant.

In agreement with others^{26–30}, we observed that M1 neurons represented the movements of the actuator even when it was passively observed by the monkey (Supplementary Fig. 7). Actuator movements (task V) replayed for the monkeys could be reconstructed from M1 activity, using a separately trained decoder (Fig. 4d), with accuracy similar to that in reconstructions made for hand control (Fig. 4c). M1 representation of the passively viewed actuator is consistent with our suggestion that a neuroprosthetic limb might become incorporated in brain circuitry¹.

Our BMBI demonstrated direct bidirectional communication between a primate brain and an external actuator. Because both the afferent and efferent channels bypassed the subject's body, we propose that BMBIs can effectively liberate a brain from the physical constraints of the body. Accordingly, future BMBIs may not be limited to limb prostheses but may include devices designed for reciprocal communication among neural structures and with a variety of external actuators.

METHODS SUMMARY

All animal procedures were performed in accordance with the National Research Council's Guide for the Care and Use of Laboratory Animals and were approved by the Duke University Institutional Animal Care and Use Committee. Two rhesus monkeys were implanted with microwire arrays in both brain hemispheres. These implants were used for both recordings and ICMS (symmetric, biphasic, charge-balanced pulse trains; 100–200 μ s, 120–200 μ A). Monkeys manipulated a joystick to cause an actuator (computer cursor or a virtual-reality arm) to reach towards up to three objects displayed on a computer monitor. The task required searching for the single object with particular artificial tactile properties. Objects consisted of a central response zone and a peripheral feedback zone. Artificial tactile feedback was delivered when the actuator entered the feedback zone and continued in the response zone. Holding the actuator over the correct object for 0.8–1.3 s produced a reward (fruit juice). Holding the actuator over an incorrect object cancelled the trial. In brain control trials, the actuator was controlled by cortical ensemble activity decoded using an unscented Kalman filter²³. An interleaved scheme of alternating recording and stimulation subintervals (50 ms each, 50% duty cycle) was implemented to achieve concurrent afferent and efferent operations. In all offline analyses, ICMS periods were excluded from calculations of neuronal firing rates. The virtual-reality arm was animated using MOTIONBUILDER (Autodesk).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 June; accepted 17 August 2011.

Published online 5 October 2011.

- Lebedev, M. A. & Nicolelis, M. A. Brain-machine interfaces: past, present and future. *Trends Neurosci.* **29**, 536–546 (2006).
- Nicolelis, M. A. & Lebedev, M. A. Principles of neural ensemble physiology underlying the operation of brain-machine interfaces. *Nature Rev. Neurosci.* **10**, 530–540 (2009).

3. Chapin, J. K., Moxon, K. A., Markowitz, R. S. & Nicolelis, M. A. L. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neurosci.* **2**, 664–670 (1999).
4. Velliste, M., Perel, S., Spalding, M. C., Whitford, A. S. & Schwartz, A. B. Cortical control of a prosthetic arm for self-feeding. *Nature* **453**, 1098–1101 (2008).
5. Moritz, C. T., Perlmutter, S. I. & Fetz, E. E. Direct control of paralysed muscles by cortical neurons. *Nature* **456**, 639–642 (2008).
6. Wessberg, J. *et al.* Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* **408**, 361–365 (2000).
7. Taylor, D. M., Helms-Tillery, S. I. & Schwartz, A. B. Direct cortical control of 3D neuroprosthetic devices. *Science* **296**, 1829–1832 (2002).
8. Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R. & Donoghue, J. P. Instant neural control of a movement signal. *Nature* **416**, 141–142 (2002).
9. Carmena, J. M. *et al.* Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* **1**, e42 (2003).
10. Johansson, R. S. & Westling, G. Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects. *Exp. Brain Res.* **56**, 550–564 (1984).
11. Flanagan, J. R. & Wing, A. M. Modulation of grip force with load force during point-to-point arm movements. *Exp. Brain Res.* **95**, 131–143 (1993).
12. James, T. W., Kim, S. & Fisher, J. S. The neural basis of haptic object processing. *Can. J. Exp. Psychol.* **61**, 219–229 (2007).
13. Chatterjee, A., Aggarwal, V., Ramos, A., Acharya, S. & Thakor, N. V. A brain-computer interface with vibrotactile biofeedback for haptic information. *J. Neuroeng. Rehabil.* **4**, 40 (2007).
14. Kaczmarek, K., Webster, J., Bach-y-Rita, P. & Tompkins, W. Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Trans. Biomed. Eng.* **38**, 1–16 (1991).
15. Marasco, P. D., Schultz, A. E. & Kuiken, T. A. Sensory capacity of reinnervated skin after redirection of amputated upper limb nerves to the chest. *Brain* **132**, 1441–1448 (2009).
16. O'Doherty, J. E., Lebedev, M. A., Hanson, T. L., Fitzsimmons, N. A. & Nicolelis, M. A. A brain-machine interface instructed by direct intracortical microstimulation. *Front. Integr. Neurosci.* **3**, 20 (2009).
17. Richer, F., Martinez, M., Robert, M., Bouvier, G. & Saint-Hilaire, J. M. Stimulation of human somatosensory cortex: tactile and body displacement perceptions in medial regions. *Exp. Brain Res.* **93**, 173–176 (1993).
18. London, B. M., Jordan, L. R., Jackson, C. R. & Miller, L. E. Electrical stimulation of the proprioceptive cortex (area 3a) used to instruct a behaving monkey. *IEEE Trans. Neural Syst. Rehabil. Eng.* **16**, 32–36 (2008).
19. Romo, R., Hernandez, A., Zainos, A. & Salinas, E. Somatosensory discrimination based on cortical microstimulation. *Nature* **392**, 387–390 (1998).
20. Fitzsimmons, N. A., Drake, W., Hanson, T. L., Lebedev, M. A. & Nicolelis, M. A. Primate reaching cued by multichannel spatiotemporal cortical microstimulation. *J. Neurosci.* **27**, 5593–5602 (2007).
21. Lederman, S. J. & Klatzky, R. L. Hand movements: a window into haptic object recognition. *Cognit. Psychol.* **19**, 342–368 (1987).
22. Lebedev, M. A. *et al.* Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain-machine interface. *J. Neurosci.* **25**, 4681–4693 (2005).
23. Li, Z. *et al.* Unscented Kalman filter for brain-machine interfaces. *PLoS ONE* **4**, e6243 (2009).
24. Lebedev, M. A., Denton, J. M. & Nelson, R. J. Vibration-entrained and premovement activity in monkey primary somatosensory cortex. *J. Neurophysiol.* **72**, 1654–1673 (1994).
25. Liu, Y., Denton, J. M. & Nelson, R. J. Neuronal activity in primary motor cortex differs when monkeys perform somatosensory and visually guided wrist movements. *Exp. Brain Res.* **167**, 571–586 (2005).
26. Cisek, P. & Kalaska, J. F. Neural correlates of mental rehearsal in dorsal premotor cortex. *Nature* **431**, 993–996 (2004).
27. Graziano, M. S., Cooke, D. F. & Taylor, C. S. Coding the location of the arm by sight. *Science* **290**, 1782–1786 (2000).
28. Maravita, A. & Iriki, A. Tools for the body (schema). *Trends Cogn. Sci.* **8**, 79–86 (2004).
29. Tkach, D., Reimer, J. & Hatsopoulos, N. G. Observation-based learning for brain-machine interfaces. *Curr. Opin. Neurobiol.* **18**, 589–594 (2008).
30. Dushanova, J. & Donoghue, J. Neurons in primary motor cortex engaged during action observation. *Eur. J. Neurosci.* **31**, 386–398 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Dimitrov for conducting the animal neurosurgeries; G. Lehw and J. Meloy for building brain implants; J. Fruh for rendering the virtual-reality monkey arm; T. Phillips, L. Oliveira and S. Halkiotis for technical support; and E. Thomson and Z. Li for comments. This research was supported by DARPA (award N66001-06-C-2019), TATRC (award W81XWH-08-2-0119), the NIH (award NS073125), NICHD/OD (award RC1HD063390) and NIH Director's Pioneer Award DP1OD006798, to M.A.L.N. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Office of the NIH Director or the NIH.

Author Contributions J.E.O'D., M.A.L. and M.A.L.N. designed experiments, analysed data and wrote the paper; J.E.O'D., M.A.L., P.J.I. and K.Z.Z. conducted experiments; and S.S. and H.B. developed the virtual-reality monkey arm.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.A.L.N. (nicoleli@neuro.duke.edu).

METHODS

Subjects and implants. Two adult rhesus macaque monkeys (*Macaca mulatta*) participated in this study. Each monkey was implanted with four 96-microwire arrays constructed of stainless steel 304. Each hemisphere received two arrays: one in the upper-limb representation area and one in the lower-limb representation area. These arrays sampled neurons in both M1 and S1. We used recordings from the right-hemisphere arm arrays in each monkey, because each manipulated the joystick with its left hand. Within each array, microwires were grouped in two four-by-four, uniformly spaced grids each consisting of 16 electrode triplets. The separation between electrode triplets was 1 mm. The electrodes in each triplet had three different lengths, increasing in 300- μ m steps. The penetration depth of each triplet was adjusted with a miniature screw. After adjustments during the month following the implantation surgery, the depth of the triplets was fixed. The longest electrode in each triplet penetrated to a depth of 2 mm as measured from the cortical surface.

Tasks. The monkeys were trained to manipulate a computer cursor or a virtual-reality arm and to reach, using this actuator, towards objects displayed on a computer monitor. The objects were visually identical, but had different tactile properties as conveyed by ICMS of S1. In hand control, each trial commenced when the monkey held the joystick with their working hand. Then a target appeared in the centre of the screen. The monkey had to hold the actuator within that centre target for a random hold time uniformly drawn from the interval 0–2 s. After this, the central target disappeared and was replaced by a set of virtual objects radially arranged about the centre of the screen. Each of these consisted of a central response zone and a peripheral feedback zone, distinguished by their shading (Supplementary Fig. 1c). Tactile feedback was delivered in the feedback zone or the corresponding response zone. For monkey M, the radius of the response zone varied from 1.5 to 4.0 cm and the radius of the feedback zone varied from 4.5 to 7.25 cm, across all tasks and sessions. For monkey N, the radius of the response zone varied from 1.5 to 4.5 cm and the radius of the feedback zone varied from 4.75 to 9.5 cm, across all tasks and sessions. A trial was concluded when the monkey placed the actuator within the response zone for a hold interval (800–1,300 ms for hand control, depending on the session; 300–500 ms for brain control) or the monkey released the joystick handle (in hand control trials). The next trial commenced after an intertrial interval of 500 ms.

The sequence of events was the same during brain control trials. In some brain control sessions, the joystick was removed from the behavioural set-up. For these, each new trial commenced following the previous intertrial interval without the requirement for the monkey to hold the joystick. In tasks I–III, monkeys chose from a set of two objects. In task I, the monkeys had to choose between RAT and NAT for fixed object locations. In task II, RAT and NAT were presented on the screen at different angular locations in each trial. In task III, object number and spatial arrangement were the same as in task II, but RAT and UAT were used. In task IV, three objects were used (RAT, UAT and NAT) and their arrangement on the screen varied from trial to trial. Finally, in task V, the virtual-reality monkey arm replaced the computer cursor.

Psychometric measurements. Psychometric measurements determined the minimum ICMS amplitude that the monkeys could discriminate (Supplementary Fig. 2b). In these measurements, the ICMS amplitude was different in every trial. In each psychometric session, a range of amplitudes was selected such that about half were in a range clearly above the monkeys' threshold for discrimination and half were in a range of unknown discriminability.

Catch trials. In some sessions, a small percentage of trials (typically 1%) were designated as catch trials. In these trials, the microstimulator delivered pulse trains with zero amplitude, but all other aspects of the behavioural task remained the same. This allowed us to confirm that there were no unintentional sources of information that the monkeys could use to perform the tasks.

Algorithms. An Nth-order unscented Kalman filter²³ (UKF) was used for brain control predictions. Up to a tenth-order UKF was used in some sessions, but in most sessions we found that the third-order UKF was sufficient. The filter parameters were fitted on the basis of either the hand movements of the monkeys while they performed the task using a joystick or on passive observation of actuator movements while the monkeys' arms were restrained.

ICMS. Symmetric, biphasic, charge-balanced pulse trains were delivered in a bipolar fashion across pairs of microwires. The channels selected had clear sensory receptive fields in the upper limb (monkey M: two pairs of microwires with synchronous pulse trains) or lower limb (monkey N: one pair of microwires). For monkey M, the anodic and cathodic phases of stimulation had a pulse width of 105 μ s; for monkey N, the pulse width was 200 μ s. The anodic and cathodic phases of the stimulation waveforms were separated by 25 μ s.

Interleaved ICMS and recordings. We implemented an interleaved scheme of alternating recording and stimulation intervals (Supplementary Fig. 2a). Our BMI had a 10-Hz update rate. That is, 100 ms of past neural data were used to make predictions about the desired state of the actuator. We broke up each 100-ms interval into two 50-ms subintervals. In the first subinterval (Rec), neural activity was recorded as usual and the measured spike count was used to estimate the firing rate for the whole 100-ms interval. The second subinterval (Stim) was reserved exclusively for delivering ICMS; all spiking activity occurring in this subinterval was discarded. Whenever the actuator was in contact with a virtual object at the start of a Stim interval, an ICMS pulse train was delivered. For RAT, nine pulses of ICMS were delivered; for UAT, 18 pulses of ICMS were delivered; and for NAT, no pulses of ICMS were delivered. The neural activity in the Stim interval was discarded even for NAT, so that there would be no bias induced by ICMS-occluded neural data.

Virtual-reality monkey arm. In task V, we introduced a novel, brain-controlled virtual-reality arm with realistic kinematic movements and spatial interactions. The control loop rate was 50 Hz, with visual refreshing at 30 Hz. The arm model was designed to depict a rhesus macaque. We presented a first-person perspective of the virtual-reality arm to the monkey, who controlled the position of the hand. Arm posture was controlled using a mixture of direct control of end effectors and inverse kinematics, constrained by the physical interdependencies of the joints.

An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor

Christiane A. Opitz^{1,2*}, Ulrike M. Litzenburger^{1,2*}, Felix Sahm³, Martina Ott^{1,2}, Isabel Tritschler⁴, Saskia Trimp⁵, Theresa Schumacher^{1,2}, Leonie Jestaedt⁶, Dieter Schrenk⁷, Michael Weller⁴, Manfred Jugold⁸, Gilles J. Guillemin⁹, Christine L. Miller¹⁰, Christian Lutz¹¹, Bernhard Radlwimmer¹², Irina Lehmann⁵, Andreas von Deimling³, Wolfgang Wick^{1,13} & Michael Platten^{1,2}

Activation of the aryl hydrocarbon receptor (AHR) by environmental xenobiotic toxic chemicals, for instance 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (dioxin), has been implicated in a variety of cellular processes such as embryogenesis, transformation, tumorigenesis and inflammation. But the identity of an endogenous ligand activating the AHR under physiological conditions in the absence of environmental toxic chemicals is still unknown. Here we identify the tryptophan (Trp) catabolite kynurenine (Kyn) as an endogenous ligand of the human AHR that is constitutively generated by human tumour cells via tryptophan-2,3-dioxygenase (TDO), a liver- and neuron-derived Trp-degrading enzyme not yet implicated in cancer biology. TDO-derived Kyn suppresses antitumour immune responses and promotes tumour-cell survival and motility through the AHR in an autocrine/paracrine fashion. The TDO-AHR pathway is active in human brain tumours and is associated with malignant progression and poor survival. Because Kyn is produced during cancer progression and inflammation in the local microenvironment in amounts sufficient for activating the human AHR, these results provide evidence for a previously unidentified pathophysiological function of the AHR with profound implications for cancer and immune biology.

Degradation of Trp by indoleamine-2,3-dioxygenases 1 and 2 (IDO1/2) in tumours and tumour-draining lymph nodes inhibits antitumour immune responses^{1–5} and is associated with a poor prognosis in various malignancies⁶. Inhibition of IDO1/2 suppresses tumour formation in animal models^{1,3} and is currently tested in phase I/II clinical trials in patients with cancer⁷. The relevance of Trp catabolism to human tumour formation and progression is, however, as yet unknown.

TDO degrades Trp to Kyn in human brain tumours

A screen of human cancer cell lines revealed constitutive degradation of Trp and release of high micromolar amounts of Kyn in brain tumour cells, namely glioma cell lines and glioma-initiating cells (GICs), but not human astrocytes (Fig. 1a). IDO1 and IDO2 did not account for the constitutive Trp catabolism in brain tumours (Supplementary Fig. 1a–e). Instead, tryptophan-2,3-dioxygenase (TDO), which is predominantly expressed in the liver (Supplementary Fig. 3a, b) and is believed to regulate systemic Trp concentrations⁶, was strongly expressed in human glioma cells (Supplementary Fig. 1b) and correlated with Kyn release (Fig. 1b). Pharmacological inhibition or knockdown of *TDO* blocked Kyn release by glioma cells, whereas knockdown of *IDO1* and *IDO2* had no effect (Fig. 1c, d and Supplementary Fig. 2a), thus confirming that TDO is the central Trp-degrading enzyme in human glioma cells. In human brain tumour specimens, TDO protein levels increased with malignancy and correlated with the proliferation index (Fig. 1e–g and Supplementary Fig. 2b–d). As described previously⁸, healthy human brain showed only weak TDO staining in neurons (Fig. 1e). TDO

expression was not confined to gliomas but was also detected in other cancer types (Supplementary Fig. 3b, c). Lower Trp concentrations were measured in the serum of patients with glioma (Fig. 1h). These may not have translated into increased Kyn levels (Fig. 1h) because Kyn is taken up by other cells and metabolized to quinolinic acid. Indeed, accumulation of quinolinic acid was detected in TDO-expressing glioblastoma tissue (Fig. 1i and Supplementary Fig. 3d).

Effects of TDO-mediated Kyn release on immune cells

Kyn suppresses allogeneic T-cell proliferation⁹. Allogeneic T-cell proliferation was inversely correlated with Kyn formation by glioma-derived TDO (Fig. 2a and Supplementary Fig. 4a, b). Knockdown of *TDO* in glioma cells (Supplementary Fig. 4c, d) restored allogeneic T-cell proliferation, and the addition of Kyn to the *TDO* knockdown cells prevented the restoration of T-cell proliferation (Fig. 2b). The proliferation of CD4⁺ and CD8⁺ T cells stimulated by the T-cell receptor was inhibited by Kyn in a concentration-dependent manner (Supplementary Fig. 4e). In addition, knockdown of *TDO* resulted in enhanced lysis of glioma cells by alloreactive peripheral blood mononuclear cells (Supplementary Fig. 4f). Finally, decreased infiltration with leukocyte common antigen (LCA)-positive and CD8⁺ immune cells was observed in sections of human glioma with high TDO expression in comparison with those with low TDO expression (Fig. 2c), indicating that Kyn formation by TDO may suppress antitumour immune responses. *In vivo* experiments in immunocompetent mice demonstrated that tumours expressing TDO grew

¹Department of Neurooncology, Neurology Clinic and National Center for Tumor Diseases University Hospital of Heidelberg, 69120 Heidelberg, Germany. ²Experimental Neuroimmunology Unit, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ³Department of Neuropathology, Institute of Pathology, University Hospital of Heidelberg and Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ⁴Department of Neurology, University Hospital Zurich, 8091 Zurich, Switzerland. ⁵Department for Environmental Immunology, Helmholtz Center for Environmental Research, 04318 Leipzig, Germany. ⁶Department of Neuroradiology, University Hospital of Heidelberg, 69120 Heidelberg, Germany. ⁷Food Chemistry and Toxicology, University of Kaiserslautern, 67663 Kaiserslautern, Germany. ⁸Small Animal Imaging Center, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ⁹Department of Pharmacology, School of Medical Sciences, University of New South Wales, Sydney, 2052 Australia. ¹⁰Department of Pediatrics, Johns Hopkins University, Baltimore, MD 21287, USA. ¹¹Heidelberg Pharma AG, 68526 Ladenburg, Germany. ¹²Division of Molecular Genetics, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ¹³Clinical Cooperation Unit Neurooncology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.

*These authors contributed equally to this work.

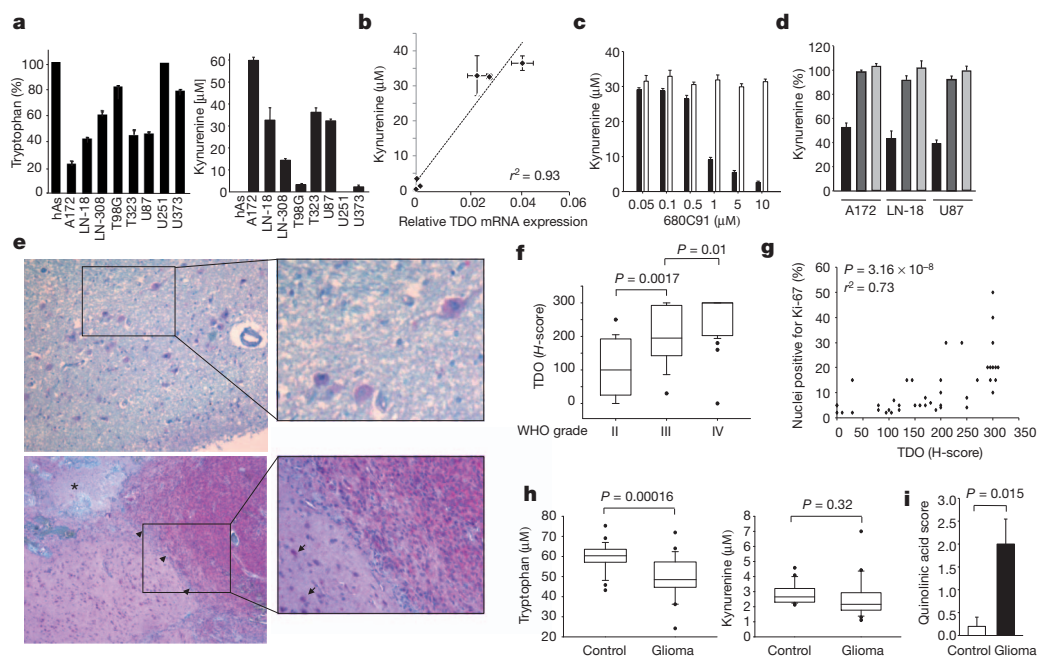


Figure 1 | TDO degrades Trp to Kyn in human brain tumours. **a**, Trp (left) and Kyn (right) content in the supernatants of human astrocytes (hAs), glioma cell lines and GICs (T323) cultured for 72 h and measured by high-performance liquid chromatography (HPLC) ($n = 4$). **b**, Correlation between *TDO* mRNA and Kyn release by human glioma cells measured by quantitative RT-PCR and HPLC ($n = 4$). **c**, Kyn concentrations in the supernatants of U87 glioma cells cultured for 48 h in the presence of the TDO inhibitor 680C91 (black bars) or its solvent (white bars; $n = 4$, $P = 0.005$, 0.002 and 0.0009 for 1, 5 and 10 μM 680C91, respectively). **d**, Kyn release by glioma cells after knockdown of *TDO* (black bars; $P = 0.000007$, 0.00007 and 0.00006, respectively), *IDO1* (dark grey bars) or *IDO2* (light grey bars) by siRNA ($n = 3$). **e**, Upper panel: weak neuronal *TDO* expression in healthy brain tissue. Lower panel: *TDO* expression in glioblastoma (WHO grade IV). *TDO* staining is in red. Asterisk indicates necrosis; arrowheads indicate the border

faster and had a higher proliferation index than *TDO*-deficient control tumours (Fig. 2d and Supplementary Fig. 4g–i). *TDO* activity in tumours suppressed antitumour immune responses *in vivo*, as demonstrated by a decreased release of interferon- γ by tumour-specific T cells and tumour cell lysis by spleen cells of mice bearing *TDO*-expressing tumours in comparison with mice bearing *TDO*-deficient tumours (Fig. 2e, f).

Effects of *TDO*-mediated Kyn release on glioma cells

We next assessed the autocrine effects of Kyn on glioma cells. Although no differences in cell cycle progression were detected between controls and glioma cells with *TDO* knockdown (Supplementary Fig. 5a), knockdown of *TDO* decreased motility and clonogenic survival (Fig. 2g, h and Supplementary Fig. 5b–d). This was mediated by Kyn because exogenous addition of Kyn restored motility and clonogenic survival in the absence of Trp (Fig. 2i, j and Supplementary Fig. 5e, f), suggesting that Kyn increases the motility of malignant glioma cells. In GICs, sphere formation was enhanced in response to Kyn (Supplementary Fig. 5g). Finally, tumour formation was impaired when *TDO* knockdown tumours were orthotopically implanted in the brains of nude mice, which are devoid of functional T cells (Fig. 2k and Supplementary Fig. 5h, i). To analyse whether *TDO*-mediated inhibition of antitumour natural killer (NK)-cell responses, which are functional in nude mice, might account for the impaired formation of *TDO* knockdown tumours, we compared subcutaneous tumour growth in the presence or absence of NK cells. NK-cell depletion (Supplementary Fig. 5j) enhanced the growth of both control and *TDO* knockout tumours but did not restore the growth of *TDO* knockout tumours to that in controls (Fig. 2l and Supplementary Fig. 5k), suggesting that Kyn generated by constitutive *TDO* activity

to infiltrated brain tissue. Insets: single tumour cells (arrows) infiltrating the adjacent brain tissue. Magnifications: $\times 40$ (main panels), $\times 400$ (upper inset) and $\times 100$ (lower inset). **f**, Plot of *TDO* expression (*H*-score; see Supplementary Methods) in brain tumours of increasing malignancy (WHO grades II–IV; grade II, $n = 18$; grade III, $n = 15$; grade IV, $n = 35$). **g**, Correlation of the Ki-67 proliferative index with the *TDO H*-score in gliomas of different WHO grades ($n = 42$). **h**, Trp (left) and Kyn (right) concentrations in the sera of 24 patients with glioblastoma and 24 age-matched and sex-matched healthy controls, measured by HPLC. **i**, Quantification of staining with quinolinic acid in healthy human brain tissue (white bar; $n = 5$) and glioblastoma tissue (black bar; $n = 5$). The data distribution in **f** and **g** is presented as box plots, showing the 25th and 75th centiles together with the median; whiskers represent the 10th and 90th centiles, respectively. Error bars indicate s.e.m.

enhances the malignant phenotype of human gliomas in an autocrine manner in the absence of functional antitumour T-cell and NK-cell responses.

Kyn activates the aryl hydrocarbon receptor

For a better understanding of the molecular mechanisms underlying the autocrine effects of Kyn on glioma cells, we performed microarray analyses of Kyn-treated glioma cells revealing broad induction of aryl hydrocarbon receptor (AHR) response genes by Kyn (Fig. 3a and Supplementary Figs 6a, b and 7). Pathway analyses showed that the 25 genes that were most strongly induced by Kyn treatment in U87 cells at 8 h and at 24 h were all directly or indirectly regulated by the AHR (Fig. 3a and Supplementary Fig. 6b). The AHR is a transcription factor of the basic helix–loop–helix (bHLH) Per–Arnt–Sim (PAS) family, which is activated by xenobiotics such as benzo[a]pyrene and 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD)¹⁰. Malignant glioma cell lines and GICs express the AHR constitutively (Supplementary Fig. 6c)¹¹, and upregulation of AHR target genes by Kyn was confirmed in two different glioma cell lines (Supplementary Fig. 6d, e). Kyn led to translocation of the AHR into the nucleus after 1 h, thus showing an immediate effect of Kyn on the AHR (Fig. 3b, c and Supplementary Fig. 8a). In accordance with this, western blot analyses of Kyn-activated tumour cells showed decreased cytoplasmic localization paralleled by increased nuclear accumulation of the AHR comparable to that induced by TCDD (Fig. 3d). In the nucleus the AHR forms a heterodimer with the AHR nuclear translocator (ARNT) that interacts with the core-binding motif of the dioxin-responsive elements (DRE) located in regulatory regions of AHR target genes¹². Kyn induced DRE-luciferase activity in glioma cells, with a concentration giving half-maximal response (EC_{50}) of 36.6 μM (Fig. 3e,

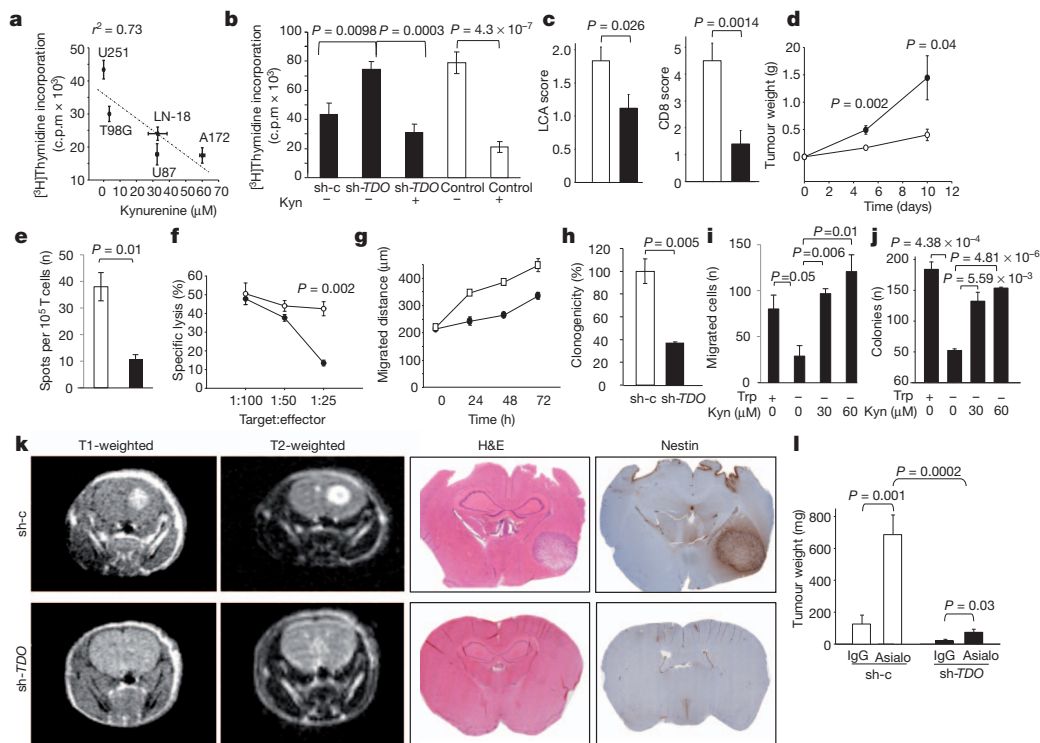


Figure 2 | Paracrine and autocrine effects of TDO-mediated Kyn release by glioma cells. **a**, Correlation of the proliferation of peripheral blood mononuclear cells (PBMCs) cultured with allogeneic glioma cell lines with the Kyn release of the glioma cells ($n = 3$). **b**, Proliferation of PBMCs cultured with allogeneic *TDO*-expressing control U87 glioma cells (sh-c) in comparison with U87 glioma cells with a stable short hairpin RNA-mediated knockdown of *TDO* (sh-TDO), with or without 100 μ M Kyn (black bars), in comparison with PBMCs alone with or without 100 μ M Kyn (white bars; $n = 3$). **c**, Quantification of LCA⁺ cells (left) and CD8⁺ cells (right) stained in human glioma sections with low *TDO* expression (*H*-score < 150, white bar; $n = 12$ for LCA, $n = 10$ for CD8) and in human glioma sections with high *TDO* expression (*H*-score ≥ 150 , black bar; $n = 17$ for LCA, $n = 10$ for CD8). **d**, Growth of *Tdo*-deficient GL261 murine glioma cells stably transfected with *Tdo* (filled circles) or empty vector (open circles) injected subcutaneously into the flank of C57BL/6N mice was monitored using metric calipers ($n = 6$). Tumour weight (g) was calculated as $0.5 \times \text{length (cm)} \times \text{width}^2 \text{ (cm}^2\text{)}$. **e**, Interferon- γ release by T cells of mice bearing subcutaneous *Tdo*-expressing tumours (black bar) in comparison with T cells of mice bearing *Tdo*-deficient

tumours (white bar) after re-stimulation with glioma lysates measured by ELISpot ($n = 3$). **f**, Lysis of GL261 murine glioma cells by spleen cells of mice with *Tdo*-expressing GL261 tumours in comparison with those with subcutaneous *Tdo*-deficient GL261 tumours, measured by chromium release ($n = 4$). **g**, Quantification of the migrated distances of sh-c (open squares) and sh-TDO (filled circles) cells into a collagen matrix ($n = 3$, $P = 0.004$, 0.0005 and 0.01 for 24, 48 and 72 h, respectively). **h**, Clonogenic survival of sh-c (white bar) and sh-TDO (black bar) U87 cells ($n = 3$). **i**, Matrigel Boyden chamber assay of U87 glioma cells in the absence or presence of 70 μ M Trp without or with 30 or 60 μ M Kyn ($n = 3$). **j**, Clonogenic survival of LN-18 glioma cells in the absence or presence of 70 μ M Trp without or with 30 or 60 μ M Kyn ($n = 3$). **k**, Representative cranial MRIs, haematoxylin/eosin staining (H&E) and nestin staining of *CD1 nul/nul* mice implanted with sh-c (upper panels) or sh-TDO (lower panels) U87 glioma cells. The images are representative of two independent experiments ($n = 6$). **l**, Tumour weight of sh-c (white bars) and sh-TDO (black bars) U87 glioma cells injected subcutaneously in the flank of *CD1 nul/nul* mice that were treated either with anti-asialo GM1 antibody (asialo) for NK-cell depletion or control IgG (IgG) ($n = 8$). Error bars indicate s.e.m.

Supplementary Fig. 8b). AHR activation was unique to Kyn in a panel of Trp catabolites (Supplementary Table 1). An ethoxyresorufin-O-deethylase (EROD) assay confirmed the induction of the functional AHR target gene *CYP1A1*, encoding cytochrome P₄₅₀, family 1, subfamily A, polypeptide 1, with an EC₅₀ of 12.3 μ M for Kyn (Supplementary Fig. 8c). Radioligand binding assays with mouse liver cytosol from *Ahr*-proficient and *Ahr*-deficient mice showed that Kyn binds to the AHR with an apparent K_d of roughly 4 μ M (Fig. 3f).

Activation of the AHR and upregulation of AHR-regulated gene expression in response to Kyn were inhibited by the AHR antagonist dimethoxyflavone or knockdown of *AHR* (Fig. 3g and Supplementary Fig. 8g–k), indicating that Kyn is a specific agonist of the AHR. The involvement of the same or similar AHR residues in the binding to Kyn, TCDD and 3-methylcholanthrene was confirmed by the fact that dimethoxyflavone inhibited the activation of the AHR by all three ligands (Supplementary Fig. 8g–i).

The effects of TDO-derived Kyn are mediated by the AHR

The endogenous production of Kyn in glioma cells was sufficient to activate the AHR, because knockdown of *TDO* decreased the expression of AHR-regulated genes (Fig. 3h and Supplementary Fig. 8l–o).

Because mean Kyn concentrations of $37.01 \pm 13.4 \mu$ M were measured in U87 xenografts ($n = 6$), sufficient Kyn concentrations to activate the AHR were also reached *in vivo*.

In accordance with activation of the AHR by TDO-derived Kyn, expression of the AHR target gene *TIPARP* in LCA⁺ immune cells was observed only in human glioma sections expressing *TDO* (Fig. 4a). To determine whether *TDO* influences antitumour immune responses through the AHR we analysed the infiltration of immune cells in human glioma sections in relation to their AHR expression. Indeed, infiltration by LCA⁺ and CD8⁺ immune cells was decreased in sections of human gliomas with high AHR expression compared with those with low AHR expression (Fig. 4b). To analyse the contribution of host AHR expression to tumour growth, we compared the growth of murine tumours with and without *Tdo* expression in *Ahr*-deficient and *Ahr*-proficient mice. The growth of *Tdo*-expressing tumours was attenuated in *Ahr*-deficient mice when compared with *Ahr*-proficient mice (Fig. 4c) indicating that AHR-mediated host effects enhance tumour growth. Staining of LCA⁺ immune cells in the tumours revealed that expression of *TDO* decreased the infiltration with LCA⁺ immune cells in *Ahr*-proficient mice but not in *Ahr*-deficient mice (Fig. 4d and Supplementary Fig. 9a), suggesting that

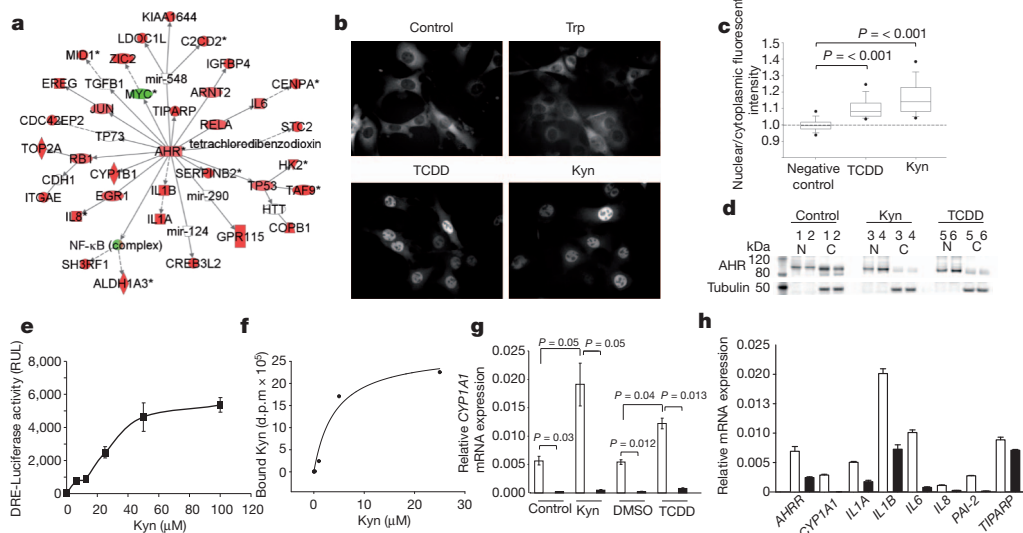


Figure 3 | Kyn activates the AHR. **a**, Connection of the 25 genes that were most strongly induced by Kyn treatment in U87 cells after 8 h to AHR signalling (red, upregulation; green, downregulation). **b**, Translocation of green fluorescent protein (GFP)-tagged AHR into the nucleus of mouse hepatoma cells, which do not degrade Trp, after 3 h treatment with 50 μ M Kyn, 50 μ M Trp or 1 nM TCDD (negative control: medium). **c**, Ratios of nuclear to cytoplasmic fluorescent intensity in cells with GFP-tagged AHR after 3 h of indicated treatment (negative control: medium; positive control: 1 nM TCDD, 50 μ M Kyn). The data distribution is represented by box plots, showing the 25th and 75th centiles together with the median; whiskers represent the 10th and 90th centiles, respectively ($P < 0.001$, one-way analysis of variance on ranks, followed by Dunn's method). **d**, AHR Western blots of two different nuclear (N) and cytoplasmic (C) fractions each of control (lanes 1 and 2), Kyn-treated (lanes 3

and 4) and TCDD-treated (lanes 5 and 6) human LN-229 glioma cells. **e**, Dioxin-responsive element (DRE) chemical activated luciferase gene expression in U87 glioma cells treated with the indicated Kyn concentrations ($n = 2$). **f**, Radioligand binding assay with indicated concentrations of 1-[3 H]Kyn using mouse liver cytosol from *Ahr*-proficient and *Ahr*-deficient mice. Specific binding was calculated by subtracting the radioactivity measured in *Ahr*-deficient cytosol from that of *Ahr*-proficient cytosol ($n = 4$). **g**, *CYP1A1* mRNA expression in sh-*AHR* LN-308 glioma cells (black bars) in comparison with controls (sh-c; white bars) treated with 100 μ M Kyn, 1 nM TCDD or controls ($n = 4$). **h**, mRNA expression of AHR target genes in sh-*TDO* (black bars) in comparison with sh-c U87 glioma cells (white bars; $n = 4$) (*AHR*, $P = 0.02$; *CYP1A1*, $P = 0.0007$; *IL1A*, $P = 0.001$; *IL1B*, $P = 0.0000006$; *IL6*, $P = 0.0047$; *IL8*, $P = 0.01$; *PAI-2*, $P = 0.0005$; *TIPARP*, $P = 0.06$). Error bars indicate s.e.m.

TDO-mediated suppression of antitumour immune responses through the AHR contributes to the host effects enhancing the growth of *Tdo*-expressing tumours. In addition, while in *Ahr*-proficient mice the expression of *Tdo* strongly enhanced tumour growth in comparison with tumours not expressing *Tdo*, the same effect was

observed in *Ahr*-deficient mice, although to a much smaller extent (Fig. 4c). Because murine glioma cells express functional AHR (Supplementary Fig. 9b–e), these results suggest that the increase in tumour growth mediated by *TDO* in *Ahr*-deficient mice is due to autocrine effects of *TDO* on the tumour cells themselves.

This notion is supported by the fact that Kyn failed to induce motility of human glioma cells after *AHR* knockdown (Fig. 4e). In addition, the increase in clonogenic survival in response to Kyn was abolished in glioma cells with a knockdown of *AHR* (Fig. 4f). Finally,

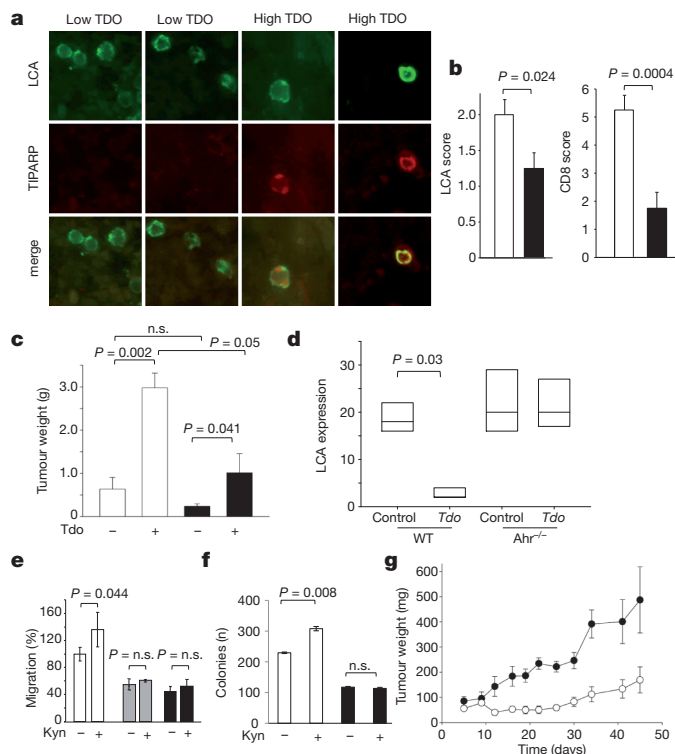


Figure 4 | The autocrine and paracrine effects of TDO-derived Kyn are mediated through the AHR. **a**, Immunofluorescence staining of LCA and TIPARP in human glioma sections with low or high *TDO* expression. Magnification $\times 400$. **b**, Quantification of LCA $^+$ cells (left) and CD8 $^+$ cells (right) stained in human glioma sections with low *AHR* expression (HistoScore < 150 , white bar; $n = 10$ for LCA, $n = 8$ for CD8) and in human glioma sections with high *AHR* expression (H-score ≥ 150 , black bar; $n = 12$ for LCA, $n = 12$ for CD8). **c**, Tumour weight measured 15 days after subcutaneous injection of murine GL261 glioma cells with and without *Tdo* expression in the flanks of *Ahr*-proficient mice (white bars) or *Ahr*-deficient mice (black bars; $n = 6$). **d**, Quantification of LCA $^+$ immune cells stained in the subcutaneous *Tdo*-proficient and *Tdo*-deficient GL261 tumours in *Ahr*-proficient and *Ahr*-deficient mice presented as box plots, showing the 25th and 75th centiles and the median ($n = 4$). WT, wild type. **e**, Migration of sh-c LN-308 glioma cells (white bars) and LN-308 glioma cells with knockdown of the *AHR* by two different shRNAs (grey bars, sh-*AHR1*; black bars, sh-*AHR2*) in the presence or absence of 100 μ M Kyn ($n = 4$). **f**, Clonogenicity of sh-c (white bars) and sh-*AHR* (black bars) LN-308 glioma cells with or without 100 μ M Kyn ($n = 3$). **g**, Growth of *AHR*-proficient (filled circles) and *AHR*-deficient (open circles) human LN-308 glioma cells injected subcutaneously into the flank of *CD1nu/nu* mice with metric callipers ($n = 7$). Tumour weight was calculated as in Fig. 2d. P values: day 5, 0.147; day 9, 0.546; day 12, 0.027; day 16, 0.008; day 19, 7.18×10^{-4} ; day 22, 1.77×10^{-5} ; day 26, 1.57×10^{-5} ; day 30, 8.49×10^{-4} ; day 34, 8.26×10^{-4} ; day 41, 0.022; day 45, 0.0477. Error bars indicate s.e.m.

in vivo experiments demonstrated that induced knockdown of AHR in human glioma cells inhibited tumour growth in immunocompromised mice (Fig. 4g and Supplementary Fig. 9f), underscoring the importance of AHR signalling for the autocrine effects of Trp degradation.

TDO-derived Kyn activates the AHR in human cancer

Next we aimed at investigating whether TDO-derived Kyn activates the AHR in human brain tumour tissue. Indeed, TDO expression correlated with the expression of the AHR and AHR target genes in human glioma tissue (Fig. 5a–c and Supplementary Fig. 9g), indicating that constitutive TDO expression in glioma cells produced sufficient Kyn levels to activate the AHR. To address whether the TDO–Kyn–AHR signalling pathway is also activated in cancers other than glioma, we analysed microarray data of diverse human tumour entities. Interestingly, TDO expression correlated with the expression of the AHR target gene *CYP1B1* not only in glioma (Fig. 5c) but also in B-cell lymphoma, Ewing sarcoma, bladder carcinoma, cervix carcinoma, colorectal carcinoma, lung carcinoma and ovarian carcinoma (Fig. 5d, Supplementary Fig. 10a and Supplementary Table 2). This finding indicates that the TDO–Kyn–AHR pathway is not confined to brain tumours but seems to be a common trait of cancers. Analysis of the Rembrandt database revealed that the overall survival of patients with glioma (WHO grades II–IV) with high expression of TDO, the AHR or the AHR target gene *CYP1B1* was reduced in comparison with patients with intermediate or low expression of these genes (Fig. 5e and Supplementary Fig. 10b)¹³. Finally, in patients with glioblastoma (WHO grade IV)¹⁴, the expression of the AHR targets *CYP1B1*, *IL1B*, *IL6* and *IL8*, which are regulated by TDO-derived Kyn in glioma cells (Fig. 3h and Supplementary Fig. 5d, e), were found to predict survival independently of WHO grade (Fig. 5f and Supplementary Fig. 10c), thus further underscoring the importance of AHR activation for the malignant phenotype of gliomas. In summary, these data suggest that endogenous tumour-derived Kyn

activates the AHR in an autocrine/paracrine fashion to promote tumour progression (Fig. 5g).

Discussion

Cancer-associated immunosuppression by Trp degradation has until now been attributed solely to the enzymatic activity of IDO in cancer cells and tumour-draining lymph nodes. IDO inhibition is therefore currently being evaluated as a therapeutic strategy in the treatment of cancer in clinical trials⁷ despite some off-target effects on human cancer cells¹⁵. We show that TDO is strongly expressed in cancer and is equally capable of producing immunosuppressive Kyn. In IDO-negative glioma cells, TDO seems to be the sole determinant of constitutive Trp degradation, indicating that TDO represents a novel therapeutic target in glioma therapy. In fact, an orally available TDO inhibitor has recently been developed¹⁶. Inhibition of TDO may not only restore antitumour immune responses but also act on the tumour cell intrinsic malignant phenotype, because we delineated the importance of constitutive Trp degradation to sustain the malignant phenotype of cancer by acting on the tumour cells themselves.

Emerging evidence indicates a tumour-promoting role of the AHR. AHR activation promotes clonogenicity and invasiveness of cancer cells^{11,17}. Transgenic mice with a constitutively active AHR spontaneously develop tumours^{18,19}, and the repressor of the AHR (AHRR) represents a tumour suppressor in multiple human cancers²⁰. The aberrant phenotype of *Ahr*-deficient mice points to the existence of endogenous AHR ligands²¹. Although different endogenously produced metabolites such as arachidonic acid metabolites, bilirubin, cyclic AMP, tryptamine and 6-formylindolo[3,2-*b*]carbazole (FICZ) have been shown to be agonists of the AHR²², their functionality has not been convincingly demonstrated in a pathophysiological context such as cancer or immune activation. The search for endogenous ligands of the AHR is therefore continuing.

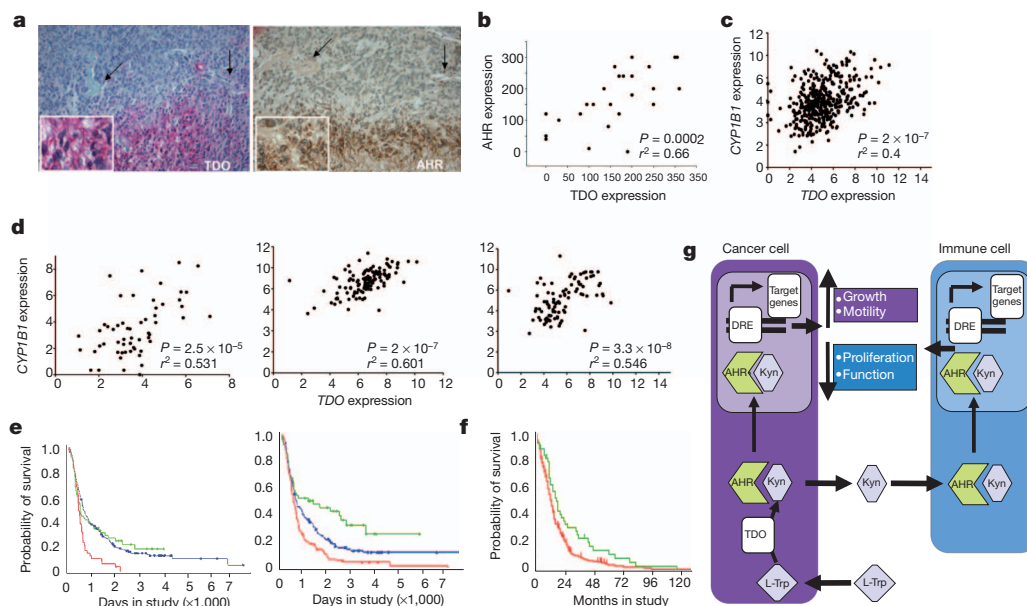


Figure 5 | TDO-derived Kyn activates the AHR in diverse human cancers, and AHR activation predicts survival in patients with glioma. **a**, Correlation of TDO expression (red) and AHR expression (brown) in consecutive sections of human glioblastoma tissue. Arrows indicate vessels for orientation. Magnification $\times 40$; insets $\times 200$. **b**, Correlation between TDO and AHR expression in human glioma tissue based on H-scores of TDO and AHR, calculated by Spearman rank correlation ($n = 26$). **c**, Correlation between TDO and *CYP1B1* expression in microarray data of human glioblastoma ($n = 396$) analysed by Spearman rank correlation. **d**, Correlation between TDO and *CYP1B1* expression in microarray data of human bladder cancer (left; $n = 58$), human lung cancer (centre; $n = 122$) and human ovarian carcinoma (right; $n = 91$) analysed by Spearman rank correlation. **e**, Survival probabilities of patients with glioma (WHO grades II–IV) with high expression (red) of TDO (left) or AHR (right) compared with those of patients with intermediate (blue) or low (green) expression of these genes derived from Rembrandt¹³. **f**, Survival probabilities of patients with glioblastoma with high expression (red) of the AHR target gene *CYP1B1* compared with those of patients with low (green) expression of *CYP1B1* derived from the glioblastoma data set of the Cancer Genome Atlas network¹⁴ ($n = 362$). **g**, Synoptic figure highlighting the autocrine and paracrine effects of TDO-derived Kyn on cancer cells and immune cells through the AHR.

g, Synoptic figure highlighting the autocrine and paracrine effects of TDO-derived Kyn on cancer cells and immune cells through the AHR.

We now link these two important pathways contributing to cancer progression by showing that Trp catabolism leads to AHR activation, and we provide evidence of a pathophysiological human condition that is associated with the production of sufficient amounts of a functionally relevant endogenous AHR ligand. Our results reveal a differential response of primary immune cells and transformed cancer cells to AHR-mediated signals, which is in line with various toxicological studies using the classical exogenous AHR ligands TCDD and 3-methylcholanthrene^{11,17,23}. Exposure to these xenobiotics leads to profound suppression of cellular and humoral immune responses²³, while also promoting carcinogenesis and inducing tumour growth^{11,17}. These cell-specific differences in AHR effects are likely to depend on the expression of factors differentially regulating AHR signal transduction such as the AHR²⁰ as well as cell-specific transcription factor crosstalk shaping the response to AHR activation²⁴.

It is likely that Kyn-mediated activation of the AHR is not only relevant in the setting of cancer. For instance, activation of the mouse and human AHR by agonistic ligands induces regulatory T cells^{25–28}. Mice with a poor-affinity AHR suffer from exacerbated autoimmune encephalomyelitis in the absence of an exogenous ligand²⁶, and Trp catabolites suppress autoimmune neuroinflammation^{29,30}, suggesting that activation of Trp catabolism represents an endogenous feedback loop to restrict inflammation through the AHR. In fact, exogenous Kyn is involved in the regulation of immune cells in mice through the AHR^{31,32}. Kyn concentrations sufficient to activate the AHR are also generated by IDO in response to inflammatory stimuli (Supplementary Fig. 11a–c). In a broader context, a significant number of malignancies arise from areas of mostly chronic infection and inflammation³³, where Trp catabolism in the tumour microenvironment is activated and sustains local immune suppression³⁴. Activation of the AHR by Kyn generated in response to inflammatory stimuli may thus constitute a previously unrecognized pathway connecting inflammation and carcinogenesis.

METHODS SUMMARY

TDO expression was analysed by immunohistochemistry in human tumours. Its relevance for Trp degradation was determined by genetic knockdown or over-expression of TDO. Trp and Kyn were measured in cell culture supernatants, human sera and xenograft tissue by high-performance liquid chromatography. Mixed leukocyte reactions, chromium release, ELISpot and staining of immune cells in tumour tissues were used to assess the immune effects of TDO activity. Cell cycle analysis, Matrigel and spheroid invasion assays, scratch assays, sphere formation assays and clonogenicity assays were employed to analyse the autocrine effects of TDO activity. All animal procedures followed the institutional laboratory animal research guidelines and were approved by the governmental authorities. Orthotopic implantation of human glioma cells with and without stable knockdown of TDO into *CD1nu/nu* mice, subcutaneous injection of these cells into NK-depleted or wild-type *CD1nu/nu* mice and subcutaneous injection of murine *Tdo*-proficient and *Tdo*-deficient GL261 cells into syngeneic C57BL/6N mice were performed to analyse the autocrine and paracrine effects of TDO activity *in vivo*. Microarray analysis of Kyn-treated human glioma cells was performed to identify signalling pathways activated by Kyn. Analysis of AHR translocation, DRE-luciferase assays and radioligand binding assays confirmed activation of the AHR by Kyn. Pharmacological inhibition and stable knockdown of the AHR (*in vitro* and *in vivo*) proved that the effects of Kyn are AHR dependent. Injection of *Tdo*-proficient and *Tdo*-deficient tumour cells into *Ahr*^{+/+} and *Ahr*^{-/-} mice was used to address the contribution of host effects to TDO-mediated cancer promotion. Finally, stainings, microarray data and clinical data on human tumour tissues were used to analyse whether TDO activates the AHR in human cancers and how this affects survival.

Received 17 November 2010; accepted 17 August 2011.

Published online 5 October 2011.

- Muller, A. J. *et al.* Inhibition of indoleamine 2,3-dioxygenase, an immunoregulatory target of the cancer suppression gene *Bin1*, potentiates cancer chemotherapy. *Nature Med.* **11**, 312–319 (2005).
- Munn, D. H. & Mellor, A. L. Indoleamine 2,3-dioxygenase and tumor-induced tolerance. *J. Clin. Invest.* **117**, 1147–1154 (2007).

- Uyttenhove, C. *et al.* Evidence for a tumoral immune resistance mechanism based on tryptophan degradation by indoleamine 2,3-dioxygenase. *Nature Med.* **9**, 1269–1274 (2003).
- Ball, H. J. *et al.* Characterization of an indoleamine 2,3-dioxygenase-like protein found in humans and mice. *Gene* **396**, 203–213 (2007).
- Metz, R. *et al.* Novel tryptophan catabolic enzyme IDO2 is the preferred biochemical target of the antitumor indoleamine 2,3-dioxygenase inhibitory compound D-1-methyl-tryptophan. *Cancer Res.* **67**, 7082–7087 (2007).
- Lob, S. *et al.* Inhibitors of indoleamine-2,3-dioxygenase for cancer therapy: can we see the wood for the trees? *Nature Rev. Cancer* **9**, 445–452 (2009).
- NewLink Genetics Corporation. IDO Inhibitor Study for Relapsed or Refractory Solid Tumors (D-1MT) <<http://clinicaltrials.gov/show/NCT00739609>> (2008).
- Miller, C. L. *et al.* Expression of the kynurenine pathway enzyme tryptophan 2,3-dioxygenase is increased in the frontal cortex of individuals with schizophrenia. *Neurobiol. Dis.* **15**, 618–629 (2004).
- Frumento, G. *et al.* Tryptophan-derived catabolites are responsible for inhibition of T and natural killer cell proliferation induced by indoleamine 2,3-dioxygenase. *J. Exp. Med.* **196**, 459–468 (2002).
- Denison, M. S. & Nagy, S. R. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu. Rev. Pharmacol. Toxicol.* **43**, 309–334 (2003).
- Gramatzki, D. *et al.* Aryl hydrocarbon receptor inhibition downregulates the TGF- β /Smad pathway in human glioblastoma cells. *Oncogene* **28**, 2593–2605 (2009).
- Reyes, H., Reisz-Porszasz, S. & Hankinson, O. Identification of the Ah receptor nuclear translocator protein (Arnt) as a component of the DNA binding form of the Ah receptor. *Science* **256**, 1193–1195 (1992).
- National Cancer Institute. REMBRANDT Home Page (<http://rembrandt.nci.nih.gov>) (2005).
- The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
- Opitz, C. A. *et al.* The indoleamine 2,3-dioxygenase (IDO) inhibitor 1-methyl-D-tryptophan upregulates IDO1 in human cancer cells. *PLoS ONE* **6**, e19823 (2011).
- Dolusic, E. *et al.* Tryptophan 2,3-dioxygenase (TDO) inhibitors. 3-(2-(pyridyl)ethyl)indoles as potential anticancer immunomodulators. *J. Med. Chem.* **54**, 5320–5334 (2011).
- Bui, L. C. *et al.* Nedd9/Hef1/Cas-L mediates the effects of environmental pollutants on cell migration and plasticity. *Oncogene* **28**, 3642–3651 (2009).
- Andersson, P. *et al.* A constitutively active dioxin/aryl hydrocarbon receptor induces stomach tumors. *Proc. Natl Acad. Sci. USA* **99**, 9990–9995 (2002).
- Moennikes, O. *et al.* A constitutively active dioxin/aryl hydrocarbon receptor promotes hepatocarcinogenesis in mice. *Cancer Res.* **64**, 4707–4710 (2004).
- Zudaire, E. *et al.* The aryl hydrocarbon receptor repressor is a putative tumor suppressor gene in multiple human cancers. *J. Clin. Invest.* **118**, 640–650 (2008).
- Fernandez-Salguero, P. *et al.* Immune system impairment and hepatic fibrosis in mice lacking the dioxin-binding Ah receptor. *Science* **268**, 722–726 (1995).
- Nguyen, L. P. & Bradfield, C. A. The search for endogenous activators of the aryl hydrocarbon receptor. *Chem. Res. Toxicol.* **21**, 102–116 (2008).
- Esser, C., Rannug, A. & Stockinger, B. The aryl hydrocarbon receptor in immunity. *Trends Immunol.* **30**, 447–454 (2009).
- Frericks, M., Burgoon, L. D., Zacharewski, T. R. & Esser, C. Promoter analysis of TCDD-inducible genes in a thymic epithelial cell line indicates the potential for cell-specific transcription factor crosstalk in the AhR response. *Toxicol. Appl. Pharmacol.* **232**, 268–279 (2008).
- Apetoh, L. *et al.* The aryl hydrocarbon receptor interacts with c-Maf to promote the differentiation of type 1 regulatory T cells induced by IL-27. *Nature Immunol.* **11**, 854–861 (2010).
- Quintana, F. J. *et al.* Control of T_H17 and T_H17 cell differentiation by the aryl hydrocarbon receptor. *Nature* **453**, 65–71 (2008).
- Quintana, F. J. *et al.* An endogenous aryl hydrocarbon receptor ligand acts on dendritic cells and T cells to suppress experimental autoimmune encephalomyelitis. *Proc. Natl Acad. Sci. USA* **107**, 20768–20773 (2010).
- Veldhoen, M. *et al.* The aryl hydrocarbon receptor links T_H17-cell-mediated autoimmunity to environmental toxins. *Nature* **453**, 106–109 (2008).
- Platten, M. *et al.* Treatment of autoimmune neuroinflammation with a synthetic tryptophan metabolite. *Science* **310**, 850–855 (2005).
- Opitz, C. A., Wick, W., Steinman, L. & Platten, M. Tryptophan degradation in autoimmune diseases. *Cell. Mol. Life Sci.* **64**, 2542–2563 (2007).
- Mezrich, J. D. *et al.* An interaction between kynurenine and the aryl hydrocarbon receptor can generate regulatory T cells. *J. Immunol.* **185**, 3190–3198 (2010).
- Nguyen, N. T. *et al.* Aryl hydrocarbon receptor negatively regulates dendritic cell immunogenicity via a kynurenine-dependent mechanism. *Proc. Natl Acad. Sci. USA* **107**, 19961–19966 (2010).
- Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).
- Muller, A. J. *et al.* Chronic inflammation that facilitates tumor progression creates local immune suppression by inducing indoleamine 2,3 dioxygenase. *Proc. Natl Acad. Sci. USA* **105**, 17073–17078 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Rauschenbach, J. Reiart, S. Koch and A. Mlitzko for technical assistance, P.-N. Pfenning for help with the invasion assays, J. Blaes for help generating genetically modified cells, T. V. Lanz and I. Oezgen for help with animal experiments, C. Esser for providing *Ahr*^{-/-} mice, M. Schwarz for providing the human DRE-luciferase construct, D. Lemke for generating GICs, K. Ochs for providing human

umbilical-vein endothelial cell cDNA, A. Hertenstein for generation of CD4⁺ and CD8⁺ T cells, M. Betts and R. B. Russel for modelling of the binding of Kyn to the AHR, R. Koch for performing quantitative reverse transcriptase polymerase chain reaction analyses, R. Tudoran for generation of GL261 cells overexpressing murine TDO, M. Deponte for help with the radioligand binding assay, D. Schemmer for collecting and banking serum samples, W. Roth for providing tissue specimens, M. Remke for suggestions regarding data analysis, and G. Hämmerling, B. Arnold and M. Feuerer for discussions. This work was supported by grants from the Helmholtz Association (VH-NG-306) to M.P., the German Research Foundation to M.P. and W.W. (Deutsche Forschungsgemeinschaft SFB 938 TP K), the Hertie Foundation to W.W. and the Helmholtz Alliance on Systems Biology to S.T. and I.L. T.S. is supported by a DKFZ PhD Program stipend. C.A.O. is supported by a Heidelberg University Medical Faculty Postdoctoral Fellowship.

Author Contributions C.A.O. and U.M.L. contributed equally to this study, designed and performed experiments, analysed data and wrote the paper. F.S. and A.D. analysed

protein expression by immunohistochemistry. I.T. cloned constructs and designed experiments. S.T. and I.L. performed nuclear translocation assays. M.O. performed animal experiments. T.S. performed immune experiments. L.J. and M.J. performed MRI scans. C.L.M. and G.J.G. provided antibodies and designed experiments. D.S. performed and analysed EROD and DRE–luciferase assays. C.L. synthesized the TDO inhibitor. M.W. and W.W. were involved in study design and data interpretation. B.R. analysed microarray data. M.P. and C.A.O. conceptualized the study, interpreted data, designed experiments and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Microarray data were deposited in the Gene Expression Omnibus repository (GEO) at www.ncbi.nlm.nih.gov/geo/ under accession number GSE25272. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.P. (m.platten@dkfz-heidelberg.de).

Inhibition of BET recruitment to chromatin as an effective treatment for MLL–fusion leukaemia

Mark A. Dawson^{1,2*}, Rab K. Prinjha^{3*}, Antje Dittman^{4*}, George Giotopoulos¹, Marcus Bantscheff⁴, Wai-In Chan¹, Samuel C. Robson², Chun-wa Chung⁵, Carsten Hopf⁴, Mikhail M. Savitski⁴, Carola Huthmacher⁴, Emma Gudgin¹, Dave Lugo³, Soren Beinke³, Trevor D. Chapman³, Emma J. Roberts³, Peter E. Soden³, Kurt R. Auger⁶, Olivier Mirguet⁷, Konstanze Doehner³, Ruud Delwel⁹, Alan K. Burnett¹⁰, Phillip Jeffrey³, Gerard Drewes⁴, Kevin Lee³, Brian J. P. Huntly^{1*} & Tony Kouzarides^{2*}

Recurrent chromosomal translocations involving the mixed lineage leukaemia (MLL) gene initiate aggressive forms of leukaemia, which are often refractory to conventional therapies¹. Many MLL-fusion partners are members of the super elongation complex (SEC), a critical regulator of transcriptional elongation, suggesting that aberrant control of this process has an important role in leukaemia induction^{2,3}. Here we use a global proteomic strategy to demonstrate that MLL fusions, as part of SEC^{2,3} and the polymerase-associated factor complex (PAF)^{4,5}, are associated with the BET family of acetyl-lysine recognizing, chromatin ‘adaptor’ proteins. These data provided the basis for therapeutic intervention in MLL-fusion leukaemia, via the displacement of the BET family of proteins from chromatin. We show that a novel small molecule inhibitor of the BET family, GSK1210151A (I-BET151), has profound efficacy against human and murine MLL-fusion leukaemic cell lines, through the induction of early cell cycle arrest and apoptosis. I-BET151 treatment in two human leukaemia cell lines with different MLL fusions alters the expression of a common set of genes whose function may account for these phenotypic changes. The mode of action of I-BET151 is, at least in part, due to the inhibition of transcription at key genes (*BCCL2*, *C-MYC* and *CDK6*) through the displacement of BRD3/4, PAFc and SEC components from chromatin. *In vivo* studies indicate that I-BET151 has significant therapeutic value, providing survival benefit in two distinct mouse models of murine MLL–AF9 and human MLL–AF4 leukaemia. Finally, the efficacy of I-BET151 against human leukaemia stem cells is demonstrated, providing further evidence of its potent therapeutic potential. These findings establish the displacement of BET proteins from chromatin as a promising epigenetic therapy for these aggressive leukaemias.

Dysregulation of chromatin modifiers is a recurrent and sentinel event in oncogenesis⁶. Therapeutic strategies that selectively alter the recruitment and/or catalytic activity of these enzymes at chromatin therefore hold great promise as targeted therapies⁶. In this regard the bromodomain and extra terminal (BET) family of proteins (BRD2, BRD3, BRD4 and BRDT) provide an ideal ‘druggable’ target, because they share a common highly conserved tandem bromodomain at their amino terminus. Selective bromodomain inhibitors that disrupt the binding of BET proteins to histones have recently been described^{7,8}; however, their true therapeutic scope remains untested.

To identify the nuclear complexes associated with ubiquitously expressed BETs (BRD2/3/4), we performed a systematic global proteomic survey. Specifically, this involved a tri-partite discovery approach (Fig. 1a). In the first approach, bead-immobilized analogues

of I-BET762 (ref. 9) were incubated with HL60 nuclear extracts and bound proteins were analysed by quantitative mass spectrometry (Supplementary Table 1). This approach identified the BET isoforms and a large number of co-purifying proteins (Supplementary Tables 1 and 2), indicating that the BET isoforms reside in many distinct protein complexes. In the second approach, immunoprecipitation analyses with selective antibodies against BRD2/3/4 were performed (Supplementary Fig. 1 and Supplementary Tables 3 and 4). This was complemented with additional immunoprecipitations using selected antibodies against complex members (‘baits’) selected from the subset of proteins that were identified in the first approach (Fig. 1b right panel, Supplementary Fig. 2 and Supplementary Table 3). In the third approach, bead-immobilized histone H4(1–21; K5acK8acK12ac) acetylated peptides were used to purify protein complexes. These data were combined to highlight a list of complexes identified in all three methods (Fig. 1b left panel, Supplementary Fig. 3 and Supplementary Table 1). Finally, specificity of the I-BET762 and histone tail matrix was further assessed by competition experiments (Fig. 1c, Supplementary Figs 4, 5 and Supplementary Table 2). This strategy enabled the direct determination of the targets of the inhibitor, and the proteins associated with the target, with subunits of protein complexes exhibiting closely matching half-maximum inhibitory concentration (IC₅₀) values¹⁰. Taken together these stringent and complementary approaches provide a high confidence global data set encompassing all known^{11–13} and several novel BET protein complexes (Fig. 1b and Supplementary Fig. 3). Among the novel complexes, we observed a prominent enrichment and dose-dependent inhibition of several components of the PAFc^{4,5} and SEC^{2,3} (Fig. 1b, c), which were confirmed by reciprocal immunoprecipitations in HL60 cells (Fig. 1b). Moreover, reciprocal immunoprecipitations in two MLL-fusion leukaemia cell lines (MV4;11 and RS4;11) confirmed the relationship of SEC with BRD4 in different cellular contexts (Fig. 1d). Together these data indicate that BRD3 and BRD4 associate with the PAFc and SEC and may function to recruit these complexes to chromatin. Given that these complexes are crucial for malignant transformation by MLL fusions^{2–5} we tested the hypothesis that displacement of BET proteins from chromatin may have a therapeutic role in these leukaemias.

To progress our studies with an optimized therapeutic agent we developed I-BET151 (Fig. 1e); a novel dimethylisoxazole template, previously undisclosed as a BET bromodomain inhibitor. It was identified and optimized to retain excellent BET target potency (Fig. 1i) and selectivity (Fig. 1h, Supplementary Figs 5–10 and Supplementary Table 5) while enhancing the *in vivo* pharmacokinetics and terminal half-life to enable prolonged *in vivo* studies (Fig. 4a and Supplementary

¹Department of Haematology, Cambridge Institute for Medical Research and Addenbrookes Hospital, University of Cambridge, Cambridge CB2 0XY, UK. ²Gurdon Institute and Department of Pathology, Tennis Court Road, Cambridge CB2 1QN, UK. ³Epinova DPU, Immuno-Inflammation Centre of Excellence for Drug Discovery, GlaxoSmithKline, Medicines Research Centre, Gunnels Wood Road, Stevenage SG1 2NY, UK. ⁴Cellzome AG Meyerhofstrasse 1, 69117 Heidelberg, Germany. ⁵Molecular Discovery Research, GlaxoSmithKline R&D, Stevenage SG1 2NY, UK. ⁶Cancer Epigenetics DPU, Oncology R&D, GlaxoSmithKline, 1250 South Collegeville Road, Collegeville, Pennsylvania 19426, USA. ⁷Lipid Metabolism Discovery Performance Unit, GSK R&D, 91951 Les Ulis Cedex, France. ⁸University Hospital of Ulm Internal Medicine III Albert-Einstein-Allee 23, 89081 Ulm, Germany. ⁹Department of Hematology, Erasmus University Medical Center, 3015 GE Rotterdam, The Netherlands. ¹⁰Department of Hematology, Cardiff University School of Medicine, Cardiff, CF14 4XN, UK.

*These authors contributed equally to this work.

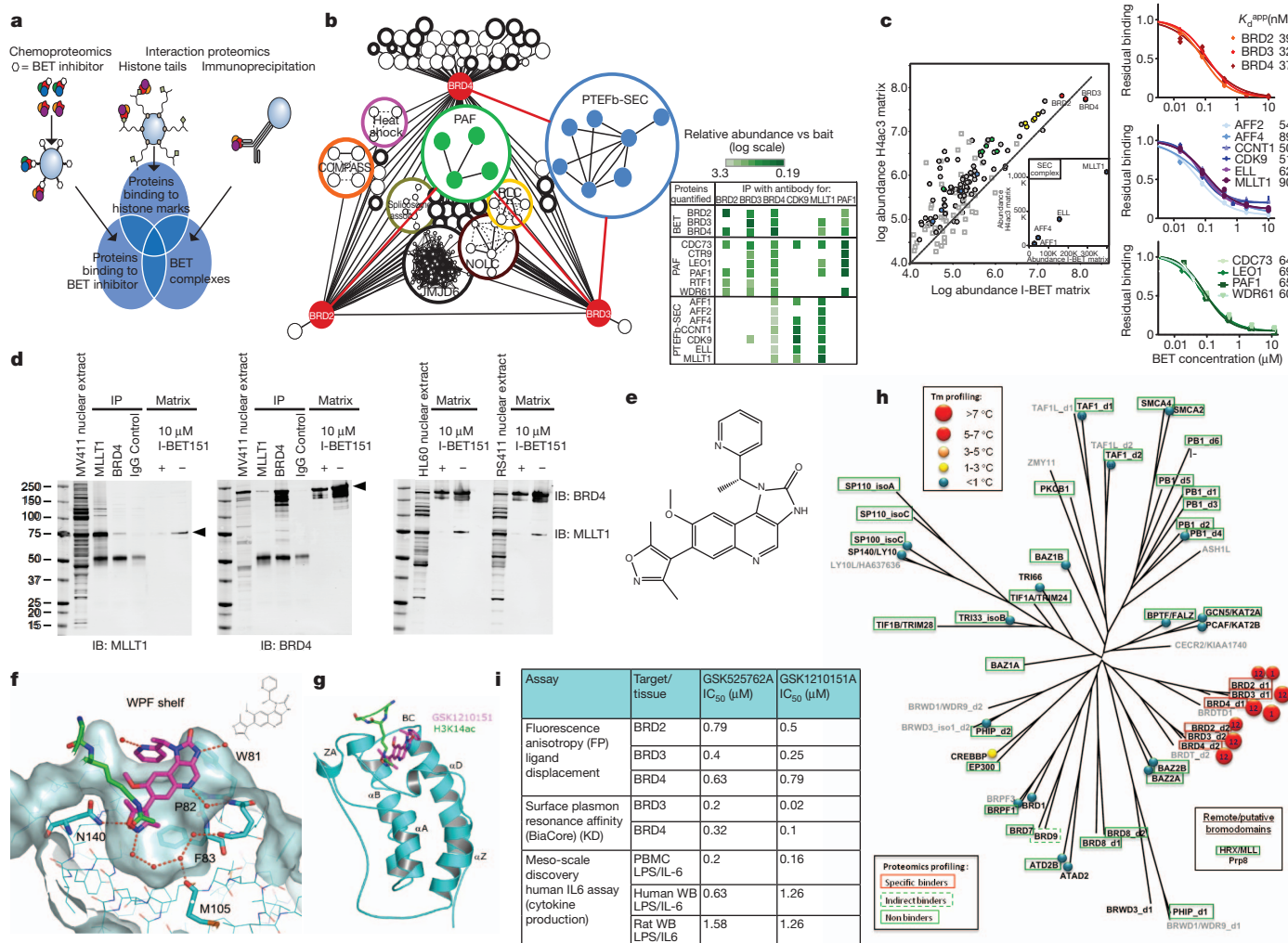


Figure 1 | A global proteomic survey identifies BET proteins as part of the PAFc and SEC. **a**, Proteomic strategy. **b**, Left, Cytoscape representation of the BET protein complex network (discussed in detail in Supplementary Fig. 3). Bold circles indicate associations confirmed by the three orthogonal methods. Right, heat map representing quantitative-mass spectrometry data following co-immunoprecipitation of BETs, PAF and SEC complex members. **c**, Differential proteomic analysis of the proteins interacting with I-BET and triple acetylated histone H4 tail. Left, affinity matrices with immobilized I-BET762 or histone H4(K5acK8acK12ac) peptide bind to the same set of BET complexes. Protein abundance was determined from signal intensities in the mass spectrometer (arbitrary units, $K = \times 1,000$). Right, competitive inhibition of the binding of BET isoforms, and SEC and PAF complex components, to the I-BET762 matrix showing matching concentration dependence. **d**, BRD4 and MLLT1 interact in HL60, MV4;11 and RS4;11 cells and binding to the I-BET762 matrix is blocked by excess I-BET151. **e**, Chemical structure of GSK1210151A (I-BET151). **f**, I-BET151 binding to the acetyl-binding pocket of BRD4-BD1 (cyan) overlaid with H3K14-acetyl peptide (green) (Protein

Database ID 3jvk). A surface representation of the BRD4-BD1 is shown with key recognition and the specificity WPF shelf identified. **g**, Ribbon representation of the BRD4-BD1 (cyan) crystal structure complexed with I-BET151 (shown in magenta stick format) overlaid with H3(12-19)K14ac peptide (green) taken from its complex with BRD4-BD1 (PDB ID 3jvk). Secondary elements of the BRD4-BD1 structure have been highlighted. **h**, Selectivity profile of I-BET-151 showing average temperature shifts (T_m) using a fluorescent thermal shift assay. Numbering inside the spheres indicates bromodomains assessed; for example, 12 signifies both bromodomains 1 and 2 have been assessed. Overlaid is the selectivity profile generated using a proteomic approach (shown as boxes around proteins, discussed in Supplementary Fig. 5). Where the bromodomains have been profiled by both thermal shift and proteomic approaches the agreement is excellent. Proteins not assessed by either technique are shown in grey. **i**, Comparison of I-BET762 and I-BET151 potency in ligand displacement assays, direct Biacore binding and lipopolysaccharide-stimulated IL-6 cytokine production from human peripheral blood mononuclear cells (PBMC) or whole blood (WB).

Fig. 20). We also generated proteomic selectivity profiles comparing I-BET151 with I-BET762 (Fig. 1h, Supplementary Fig. 5 and Supplementary Table 6). We bead-immobilized a combination of differentially acetylated histone tail peptides (Supplementary Table 7), which captured a total of 27 bromodomain proteins from HL60 nuclear extracts. Competition with excess I-BET151 or I-BET762 blocked the capture of BRD2, BRD3, BRD4, and BRD9 but had no effect on the 23 other bromodomain proteins including MLL. The inhibition of BRD9 is likely to be indirect as this protein forms a complex with BRD4 (Supplementary Table 3). Finally, a high-resolution (1.5 Å) crystal structure of I-BET151 bound to BRD4-bromodomain 1

(BD1) revealed binding to the acetylated-lysine (AcK) recognition pocket of the BET protein (Fig. 1f, g and Supplementary Fig. 10).

To assess the therapeutic efficacy and selectivity of I-BET151, we tested a panel of leukaemic cell lines harbouring a spectrum of distinct oncogenic drivers. These data demonstrated that I-BET151 has potent efficacy against cell lines harbouring different MLL-fusions (Fig. 2a and Supplementary Fig. 11). To extend these data we tested the clonogenic potential of human leukaemic cells grown in cytokine-supplemented methylcellulose containing dimethylsulphoxide (DMSO; vehicle) or I-BET151. Consistent with the profound effects in liquid culture, the colony-forming potential of MLL-fusion-driven

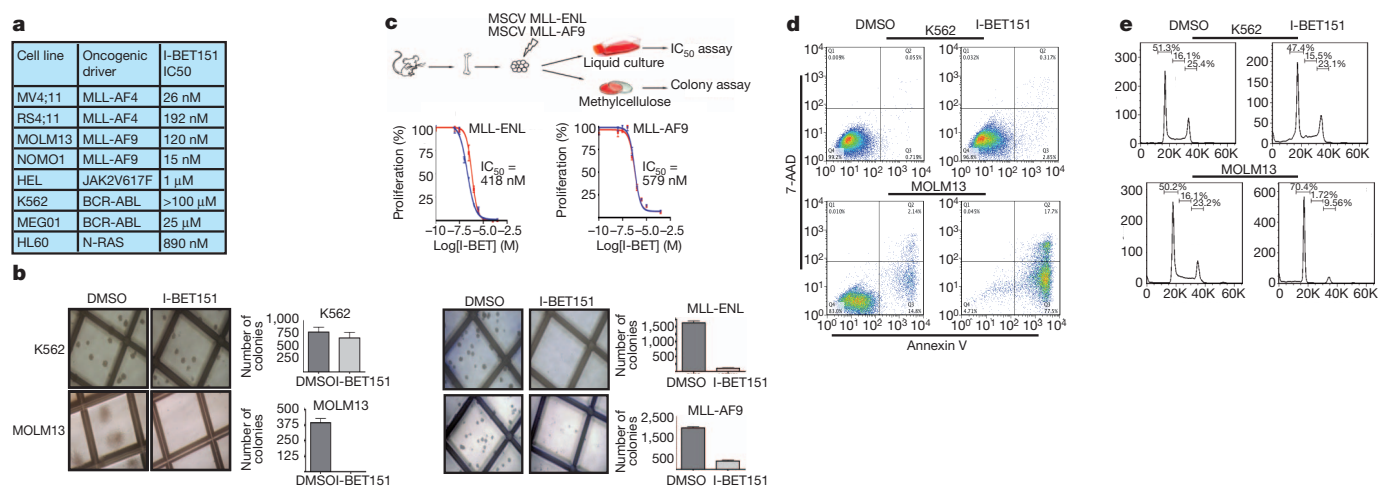


Figure 2 | I-BET151 selectively and potently inhibits MLL-fusion leukaemic cell lines *in vitro*. **a**, Human leukaemia cell lines tested using I-BET151. **b**, Clonogenic assays performed in the presence of DMSO or I-BET151. **c**, Haematopoietic progenitors were isolated from mouse bone marrow and retrovirally transformed with MLL-ENL or MLL-AF9. These cells were used in both proliferation and clonogenic assays **d**, Apoptosis was assessed by FACS

analysis after 72 h incubation with DMSO or I-BET151. **e**, Cell cycle progression was assessed by FACS analysis 24 h after incubation with DMSO or I-BET151 (y axis event count, x axis arbitrary fluorescence units). Bar graphs are represented as the mean and error bars reflect standard deviation of results derived from triplicate experiments.

leukaemias (MOLM13) was completely ablated by I-BET151, whereas leukaemias driven by tyrosine kinase activation (K562) were unaffected (Fig. 2b). In addition to the data with human leukaemic cell lines, we also confirmed the potent efficacy of I-BET151 in both liquid culture and clonogenic assays using primary murine progenitors

retrovirally transformed with either MLL-ENL or MLL-AF9 (Fig. 2c).

To investigate the mechanism of action for I-BET151, we performed fluorescence-activated cell sorting (FACS) analysis to assess apoptosis and cell cycle progression after I-BET151 treatment. Figure 2d–e and Supplementary Fig. 12 show a marked induction of apoptosis and a

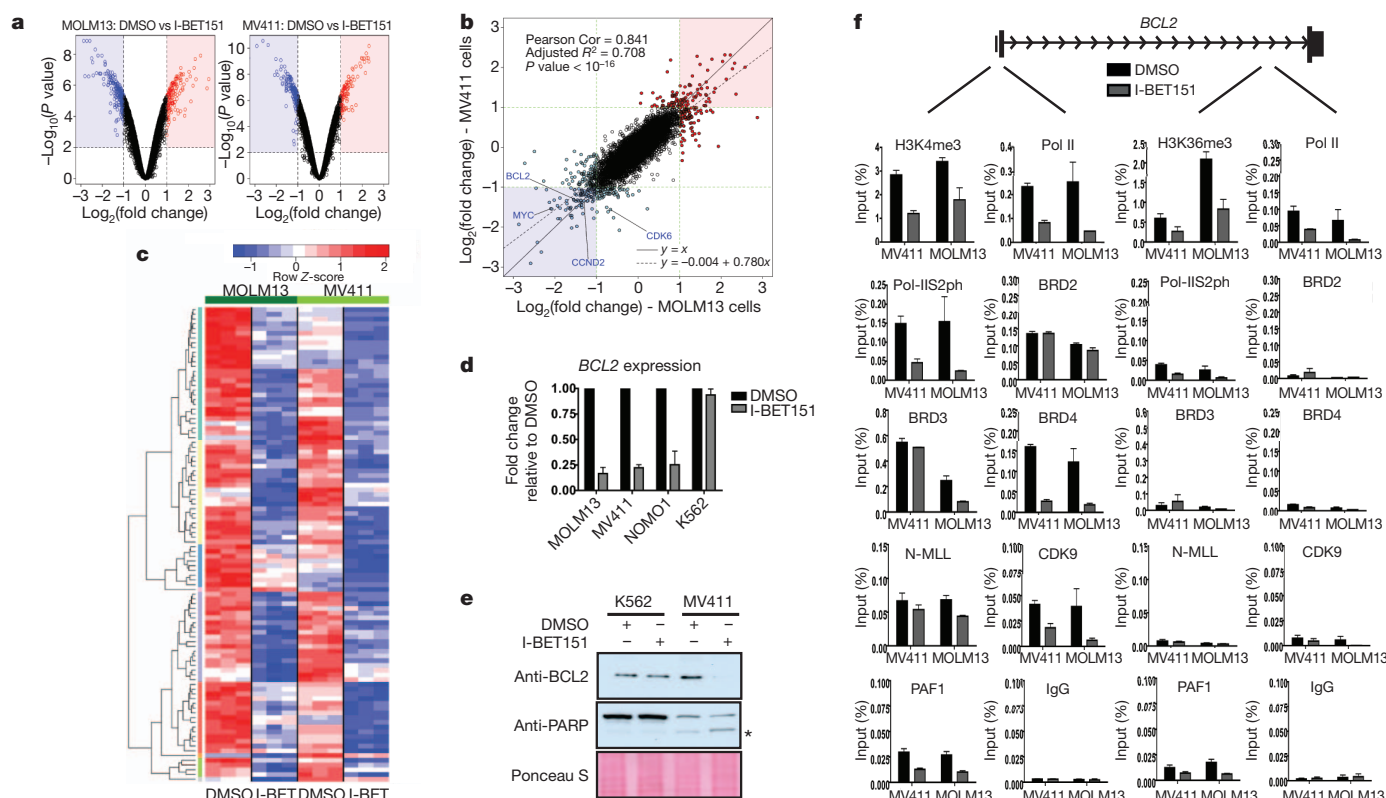


Figure 3 | Transcriptome and ChIP analyses provide mechanistic insights for the efficacy of I-BET151. **a**, Volcano plots for DMSO against I-BET151 treated samples, showing the adjusted significance P value (\log_{10}) versus fold change (\log_2). **b**, Correlation of \log_2 fold change between MV411 and MOLM13 across all genes. No genes show opposing expression changes. Lines represent the identity line (black solid), the line of best fit (black dotted), or \log_2 fold-change threshold values (green dotted). **c**, Heat map of top 100 genes

downregulated following treatment with I-BET151. **d**, $BCL2$ gene expression (normalized to $B2M$ expression) is shown. Expression level of $BCL2$ in DMSO was assigned a value of 1. **e**, Immunoblotting demonstrating a decrease in $BCL2$ and an increase in cleaved PARP (*) after I-BET151 treatment. **f**, ChIP analysis at the TSS and 3' end of $BCL2$ is illustrated. Bar graphs are represented as the mean enrichment relative to input and error bars reflect standard deviation of results derived from biological triplicate experiments.

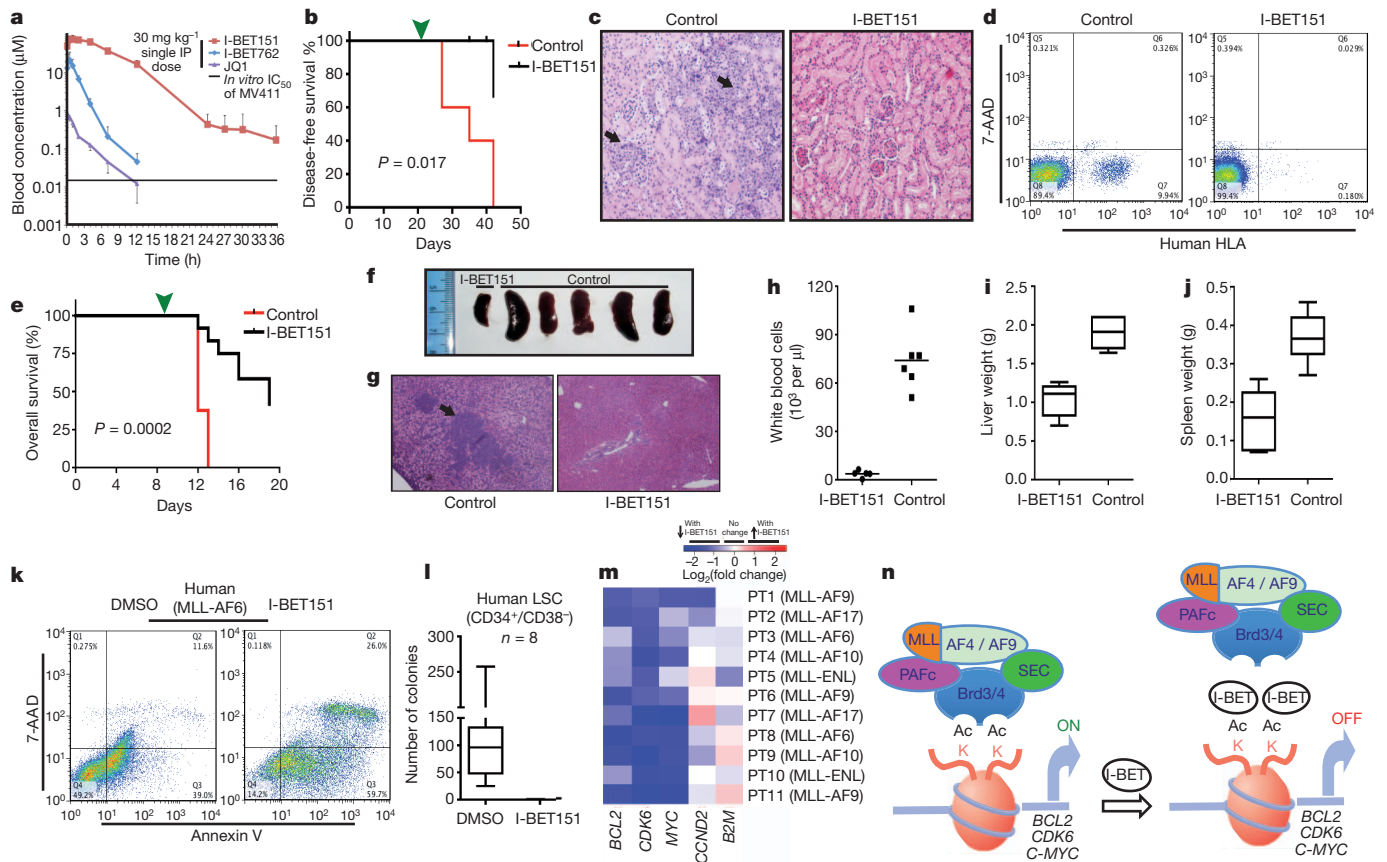


Figure 4 | I-BET151 is efficacious in *in vivo* murine models and primary patient samples of MLL-fusion leukaemia. **a**, Murine pharmacokinetic studies (mean \pm s.d., $n = 4$ per compound) comparing the blood concentration of I-BET151 with I-BET762 and JQ1. **b**, Kaplan-Meier curve of control and treated NOD-SCID mice transplanted with 1×10^7 MV4;11 cells. Green arrowhead, treatment commencement on day 21. **c**, Haematoxylin and eosin-stained histological sections of the renal parenchyma of control and treated mice. Black arrows highlight leukaemic infiltration. **d**, Representative FACS analysis from the peripheral blood of control or I-BET151-treated mice. **e**, Kaplan-Meier curve of control and treated C57BL/6 mice transplanted with 2.5×10^6 syngeneic MLL-AF9 leukaemic cells. Green arrowhead, treatment commencement on day 9. **f**, Photomicrograph of the spleen size from 5/8 control and 1/12 I-BET151-treated mice that died on day 12. **g**, Haematoxylin

and eosin-stained histological sections of the liver parenchyma from control and I-BET151-treated mice demonstrating reduced disease burden in the treated animal. **h-j**, Peripheral blood white cell count (**h**), liver weight (**i**) and spleen weights (**j**) from all the control and treated mice at the time of necropsy. **k**, Representative FACS analysis assessing apoptosis from a patient with MLL-AF6 leukaemia. **l**, Clonogenic assays with human MLL-fusion LSC isolated by FACS sorting ($\text{CD}34^+/\text{CD}38^-$) and plated in the presence of DMSO or I-BET151. **m**, Gene expression changes in human MLL-fusion leukaemia cells following treatment with I-BET151 or DMSO. The log₂ fold change in the expression level for all genes (expression level with I-BET151 treatment/expression level with DMSO) is represented. **n**, Schematic model proposing the mode of action for I-BET151 in MLL-fusion leukaemia.

prominent G₀/G₁ arrest in two MLL-fusion cell lines driven by distinct MLL fusions (MOLM13 and MV4;11 containing MLL-AF9 and MLL-AF4, respectively). In contrast, the cell cycle characteristics and apoptotic rate of K562 cells were largely unaffected at this time. These data indicate that I-BET151 alters the transcriptional programmes regulating apoptosis and cell-cycle progression in MLL-fusion leukaemias. To identify the precise transcriptional pathways controlled by I-BET151, global gene-expression analysis was performed in MOLM13 and MV4;11 cells after treatment with I-BET151 or DMSO for 6 h. This strategy allowed us to identify early I-BET151-responsive genes, before any discernable phenotypic alteration in cell cycle or apoptosis (Supplementary Fig. 12). As demonstrated previously⁷, we observed differential expression of a selective subset of genes (Fig. 3a), rather than global transcriptional dysregulation. Remarkably, the transcriptional programmes altered in the two MLL-fusion cell lines were highly correlated (Fig. 3b) and gene set enrichment analysis documented significant overlap with published MLL fusion signatures including MLL-fusion leukaemia stem cells (LSC)^{14,15} (Supplementary Fig. 13). These data are consistent with the notion that MLL fusions aberrantly co-opt the SEC and PAFc to regulate similar transcriptional

programmes. Notably, the top 100 genes concomitantly decreased in both MOLM13 and MV4;11 (Fig. 3c) contained several previously reported direct MLL targets, such as *BCL2*, *CDK6* and *MYC*, the down-regulation of which was consistent with the phenotypic consequences of I-BET151 treatment. *BCL2* is a key antiapoptotic gene implicated in the pathogenesis of MLL-fusion leukaemias^{16,17}. Consistent with these data, I-BET151 reduced the expression of *BCL2* in a third MLL-fusion cell line (NOMO1) but not in the unresponsive K562 cells (Fig. 3d), and induction of apoptosis coincided with a marked reduction in *BCL2* protein expression (Fig. 3e). Moreover, overexpression of *BCL2* in the presence of I-BET151 rescued the apoptotic phenotype (Supplementary Fig. 14). Chromatin immunoprecipitation (ChIP) analyses at the *BCL2* locus showed that 6 h of I-BET151 treatment selectively decreased the recruitment of BRD3/4 and impaired recruitment of CDK9 and PAF1 (part of SEC and PAFc, respectively) to the transcriptional start site (TSS). This correlated with reduced phosphorylation of RNA polymerase II (Pol II) on serine 2 of its carboxy-terminal domain (Pol-II-S2ph) (Fig. 3f). A similar pattern was observed at two other MLL target genes (*MYC* and *CDK6*), but not at housekeeping genes (*B2M*) whose expression was unaltered by I-BET151 (Supplementary Fig. 15).

programmes. Notably, the top 100 genes concomitantly decreased in both MOLM13 and MV4;11 (Fig. 3c) contained several previously reported direct MLL targets, such as *BCL2*, *CDK6* and *MYC*, the down-regulation of which was consistent with the phenotypic consequences of I-BET151 treatment.

BCL2 is a key antiapoptotic gene implicated in the pathogenesis of MLL-fusion leukaemias^{16,17}. Consistent with these data, I-BET151 reduced the expression of *BCL2* in a third MLL-fusion cell line (NOMO1) but not in the unresponsive K562 cells (Fig. 3d), and induction of apoptosis coincided with a marked reduction in *BCL2* protein expression (Fig. 3e). Moreover, overexpression of *BCL2* in the presence of I-BET151 rescued the apoptotic phenotype (Supplementary Fig. 14). Chromatin immunoprecipitation (ChIP) analyses at the *BCL2* locus showed that 6 h of I-BET151 treatment selectively decreased the recruitment of BRD3/4 and impaired recruitment of CDK9 and PAF1 (part of SEC and PAFc, respectively) to the transcriptional start site (TSS). This correlated with reduced phosphorylation of RNA polymerase II (Pol II) on serine 2 of its carboxy-terminal domain (Pol-II-S2ph) (Fig. 3f). A similar pattern was observed at two other MLL target genes (*MYC* and *CDK6*), but not at housekeeping genes (*B2M*) whose expression was unaltered by I-BET151 (Supplementary Fig. 15).

Together, these data indicate that the mechanism of efficacy for I-BET151 involves a selective abrogation of BRD3/4 recruitment to chromatin. The consequence of this is the inefficient phosphorylation/recruitment of Pol II. Further investigation is necessary to distinguish whether Pol II recruitment and/or elongation is primarily affected by I-BET151.

We next sought to establish the therapeutic potential of I-BET151 *in vivo*. We first characterized the pharmacokinetic properties of I-BET151 in several preclinical species (Supplementary Fig. 20) and also compared it to published inhibitors^{7,8} (Fig. 4a). We then assessed the efficacy of I-BET151 in two established models of MLL leukaemia. Our first model was a xenotransplant model of disseminated human MLL–AF4 leukaemia¹⁸. I-BET151 was delivered daily at 30 mg kg^{−1} by intraperitoneal injection from day 21 (ref. 18), and mice were humanely killed if clinical disease dictated or if there was a sequential rise in peripheral blood disease. At the experimental end-point all the control mice had succumbed to fulminant or progressive disease whereas only one out of five mice in the treated cohort had evidence of disease at low levels (Fig. 4b–d and Supplementary Fig. 16). In our second syngeneic model of murine MLL–AF9 leukaemia, 2.5 × 10⁶ leukaemic cells, established from serial transplantation, were injected into tertiary recipients. Despite the latency being reduced to less than 15 days, we waited to initiate treatment from day 9 to test the efficacy of I-BET151 in the setting of overwhelming established disease (Fig. 4e), the scenario often encountered in clinical practice. Even here I-BET151 provided a clear and marked survival benefit (Fig. 4e–j and Supplementary Fig. 17). Taken together, these data demonstrate that I-BET151 provides excellent control of MLL leukaemia progression in two distinct and complementary murine models.

Finally, to demonstrate the applicability of our findings to human disease, we tested the efficacy of I-BET151 in leukaemia cells isolated from patients with various MLL fusions. These data show that I-BET151 accelerates apoptosis (Fig. 4k and Supplementary Fig. 18), and abrogates clonogenic efficiency in bulk leukaemia (Supplementary Fig. 19) as well as isolated LSC (Fig. 4l). These effects are driven, at least in part, by downregulation of a similar transcription programme identified in MLL-fusion cell lines (Fig. 4m). Taken together, these data provide compelling evidence of therapeutic potential and suggest that disease eradication is possible.

The paradigm for epigenetic drug discovery shown here highlights an emerging role for targeting aberrant transcriptional elongation in oncogenesis^{2–5} and provides the first example in epigenetic therapy where mechanistic insights have driven targeted drug discovery and application (Fig. 4n). Together, our results suggest that perturbing the interaction of BET proteins with chromatin using I-BET151 may be of great therapeutic value in human MLL-fusion leukaemias. Using a complementary strategy and a different BET inhibitor, a separate study published in this issue concurs with this view¹⁹. Moreover, the extensive proteomic resource provided here has identified other important disease-associated proteins binding to BET proteins, such as MMSET (WHSC1), which is implicated in multiple myeloma²⁰. This raises the possibility that BET inhibitors may have an even wider therapeutic scope in oncology and perhaps in other areas of unmet need within the clinical arena.

METHODS SUMMARY

Cell culture, gene expression, chromatin immunoprecipitation and FACS analysis were performed as previously described²¹. Proteomic profiling and characterization of inhibitor specificity was performed using methodology previously described^{7,9,10}. Detailed information about the reagents and methodology used in this study is available in Supplementary Information.

Received 29 June; accepted 30 August 2011.

Published online 2 October 2011.

1. Krivtsov, A. V. & Armstrong, S. A. MLL translocations, histone modifications and leukaemia stem-cell development. *Nature Rev. Cancer* **7**, 823–833 (2007).

2. Lin, C. *et al.* AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol. Cell* **37**, 429–437 (2010).
3. Yokoyama, A., Lin, M., Nares, A., Kitabayashi, I. & Cleary, M. L. A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. *Cancer Cell* **17**, 198–212 (2010).
4. Milne, T. A. *et al.* Multiple interactions recruit MLL1 and MLL1 fusion proteins to the HOXA9 locus in leukemogenesis. *Mol. Cell* **38**, 853–863 (2010).
5. Muntean, A. G. *et al.* The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. *Cancer Cell* **17**, 609–621 (2010).
6. Rodríguez-Paredes, M. & Esteller, M. Cancer epigenetics reaches mainstream oncology. *Nature Med.* **17**, 330–339 (2011).
7. Nicodeme, E. *et al.* Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119–1123 (2010).
8. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
9. Chung, C. W. *et al.* Discovery and characterization of small molecule inhibitors of the BET family bromodomains. *J. Med. Chem.* **54**, 3827–3838 (2011).
10. Bantscheff, M. *et al.* Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nature Biotechnol.* **29**, 255–265 (2011).
11. Jang, M. K. *et al.* The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* **19**, 523–534 (2005).
12. Maruyama, T. *et al.* A mammalian bromodomain protein, Brd4, interacts with replication factor C and inhibits progression to S phase. *Mol. Cell. Biol.* **22**, 6509–6520 (2002).
13. Yang, Z. *et al.* Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol. Cell* **19**, 535–545 (2005).
14. Somerville, T. C. *et al.* Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell* **4**, 129–140 (2009).
15. Wang, J. *et al.* Conditional MLL-CBP targets GMP and models therapy-related myeloproliferative disease. *EMBO J.* **24**, 368–381 (2005).
16. Robinson, B. W. *et al.* Abundant anti-apoptotic BCL-2 is a molecular target in leukaemias with t(4;11) translocation. *Br. J. Haematol.* **141**, 827–839 (2008).
17. Wang, Q. F. *et al.* MLL fusion proteins preferentially regulate a subset of wild-type MLL target genes in the leukemic genome. *Blood* **117**, 6895–6905 (2011).
18. O'Farrell, A. M. *et al.* SU11248 is a novel FLT3 tyrosine kinase inhibitor with potent activity *in vitro* and *in vivo*. *Blood* **101**, 3597–3605 (2003).
19. Zuber, J. *et al.* RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* doi:10.1038/nature10334 (this issue).
20. Martínez-García, E. *et al.* The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. *Blood* **117**, 211–220 (2011).
21. Dawson, M. A. *et al.* JAK2 phosphorylates histone H3Y41 and excludes HP1α from chromatin. *Nature* **461**, 819–822 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. J. Dawson, A. Bannister, S. Anand and all members of the Huntly and Kouzarides laboratories. We are grateful to H. Doehner, the NCRI AML trials biobank and A. Giles for the provision of patient samples. We acknowledge D. Huang for the BCL2 expression plasmid, L. Gordon for supplying fluorescence resonance energy transfer data and R. Woodward, C. Delves, E. Jones and P. Holmes for protein production. J. Witherington, N. Parr, S. Baddeley and J. Seal provided compound selectivity data. We thank N. Deeks and L. Cutler for providing sample and PK data analysis. We acknowledge K. Smitheman and A. Wyce for help with the cellular analysis of the BET inhibitors, P. Grandi for suggestions and discussion, S. Chan for biophysical assay data, and members of the Epinova team for discussion and suggestions. We thank staff at the ESRF at Grenoble for beamline assistance. We thank T. Werner for assistance with mass spectrometry experiments and data analysis, and the members of the Cellzome Biochemistry, Mass Spectrometry, and IT teams for outstanding expertise and diligence. This work was supported by a Wellcome-Beit Intermediate Clinical Fellowship to M.A.D. The Huntly lab is funded by the Medical Research Council (UK), Leukaemia Lymphoma Research (UK), the Wellcome Trust, The Leukemia & Lymphoma Society of America, Cancer Research UK (CRUK) and the NIHR Cambridge Biomedical Research Centre. This work in the Kouzarides laboratory was funded by a programme grant from Cancer Research UK (CRUK).

Author Contributions M.A.D., R.K.P., A.D., G.D., K.L., P.J., B.J.P.H. and T.K. designed the research, interpreted data and wrote the manuscript. M.A.D., A.D., G.G., M.B., W.-I.C., S.C.R., C.-W.C., C.H., M.M.S., C.H., E.G., D.L., S.B., T.D.C., E.J.R., P.E.S., K.R.A. and O.M. performed experiments and analysed data. K.D., R.D. and A.K.B. provided patient samples. M.A.D., R.K.P. and A.D. are joint first authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. I-BET151 compound requests should be directed to K.L. (kevin.2.lee@gsk.com). Correspondence and requests for materials should be addressed to B.J.P.H. (bjph2@cam.ac.uk), M.A.D. (maf2d@cam.ac.uk) or T.K. (t.kouzarides@gurdon.cam.ac.uk).

Ocean-like water in the Jupiter-family comet 103P/Hartley 2

Paul Hartogh¹, Dariusz C. Lis², Dominique Bockelée-Morvan³, Miguel de Val-Borro¹, Nicolas Biver³, Michael Küppers⁴, Martin Emprechtinger², Edwin A. Bergin⁵, Jacques Crovisier³, Miriam Rengel¹, Raphael Moreno³, Slawomira Szutowicz⁶ & Geoffrey A. Blake²

For decades, the source of Earth's volatiles, especially water with a deuterium-to-hydrogen ratio (D/H) of $(1.558 \pm 0.001) \times 10^{-4}$, has been a subject of debate. The similarity of Earth's bulk composition to that of meteorites known as enstatite chondrites¹ suggests a dry proto-Earth² with subsequent delivery of volatiles³ by local accretion⁴ or impacts of asteroids or comets^{5,6}. Previous measurements in six comets from the Oort cloud yielded a mean D/H ratio of $(2.96 \pm 0.25) \times 10^{-4}$. The D/H value in carbonaceous chondrites, $(1.4 \pm 0.1) \times 10^{-4}$, together with dynamical simulations, led to models in which asteroids were the main source of Earth's water⁷, with ≤ 10 per cent being delivered by comets. Here we report that the D/H ratio in the Jupiter-family comet 103P/Hartley 2, which originated in the Kuiper belt, is $(1.61 \pm 0.24) \times 10^{-4}$. This result substantially expands the reservoir of Earth ocean-like water to include some comets, and is consistent with the emerging picture of a complex dynamical evolution of the early Solar System^{8,9}.

On 17 November 2010, using the Herschel Space Observatory, we determined the D/H ratio in a comet from a reservoir other than the Oort cloud—103P/Hartley 2. Such Jupiter-family comets are believed to originate from the Kuiper belt, which exists beyond the orbits of the giant planets at radii between 30 and 50 astronomical units¹⁰ (1 AU is the average Earth–Sun distance). In contrast, Oort-cloud comets are theorized to have originated from radii near the gas giants and to have been subsequently ejected to the Oort cloud ($>5,000$ AU)¹¹. The Herschel measurement therefore traces the water D/H ratio in a new population of water-ice-rich bodies in the Solar System that are a potential source of water on the Earth.

To obtain an accurate determination of the D/H ratio in water, we carried out simultaneous observations of optically thin isotopic variants of water, specifically HDO and H₂¹⁸O (Fig. 1), as part of our Solar System observing programme¹². This was critical for comet 103P/Hartley 2, whose activity and water outgassing rates exhibited significant short-term variations¹³. We used state-of-the-art excitation models to determine the HDO and H₂¹⁸O beam integrated column densities and production rates from the measured line intensities. Observation and modelling details are given in Supplementary Information. A critical point is that all observations sampled the same region of the coma, about 6,500 km in diameter.

The retrieved gas column densities and production rates are sensitive to collisional cross-sections, along with the density and temperature profiles of H₂O and electrons, and we thus considered a range of model parameters (Table 1). Although the production rates determined for the various model parameters differ slightly, the value of the D/H ratio is estimated to be $(1.61 \pm 0.24) \times 10^{-4}$. In our analysis, we assumed an H₂¹⁶O/H₂¹⁸O ratio of 500 ± 50 , a range that encompasses the Earth value and is consistent with previous measurements in cometary water¹⁴ (see also Supplementary Information). The quoted 1 σ uncertainty in the D/H ratio includes a 5% uncertainty related to modelling.

Our measured D/H value is substantially larger than that which characterized the young Sun (4.5 Gyr ago; the protosolar ratio), believed to be about 2.1×10^{-5} , which in turn is slightly higher than the value found in the local interstellar medium today (1.6×10^{-5}) and

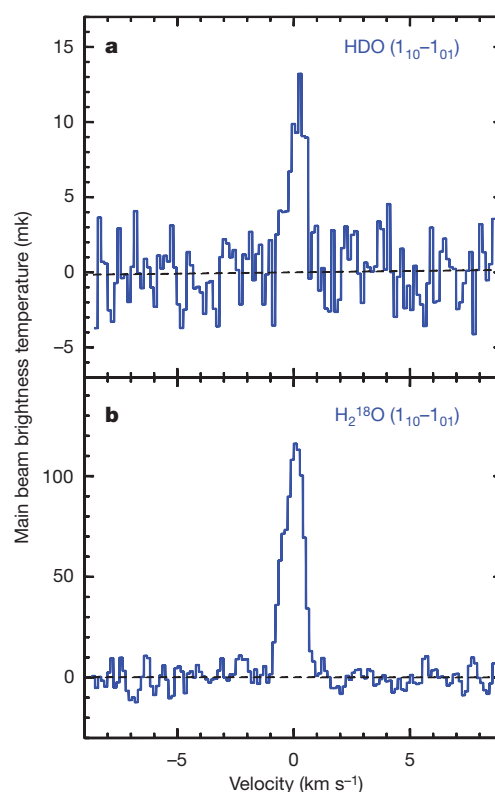


Figure 1 | Submillimetre water emission lines from comet 103P/Hartley 2. The time of the observations was 20 days after perihelion, when the comet was 1.095 AU from the Sun and 0.212 AU from Herschel. Because the H₂O ground state rotational lines in comets are optically thick^{29,30}, observations of the rare oxygen isotopic counterpart, H₂¹⁸O, provide a more reliable reference for the D/H determination. The spectra of the 1₁₀–1₀₁ lines of HDO (a) and H₂¹⁸O (b) at 509.292 and 547.676 GHz, respectively, were obtained with the Heterodyne Instrument for the Far Infrared (HIFI) High Resolution Spectrometer (HRS) between 17.28 and 17.64 November 2010 UT. The line intensities, expressed in the main-beam brightness temperature scale, are 0.011 ± 0.001 and 0.117 ± 0.002 K km s^{−1}, for HDO and H₂¹⁸O respectively, averaging the two instrument polarizations. The velocity scale is given relative to the velocity of the comet's nucleus. The spectral resolution is 141 and 132 m s^{−1} for the HDO and H₂¹⁸O spectra, respectively. For details of the observational sequence and basic parameters of the data analysis, see Supplementary information.

¹Max-Planck-Institut für Sonnensystemforschung, Max-Planck-Str. 2, 37191 Katlenburg-Lindau, Germany. ²California Institute of Technology, Pasadena, California 91125, USA. ³LESIA-Observatoire de Paris, CNRS, UPMC, Université Paris-Diderot, 5 place Jules Janssen, 92195 Meudon, France. ⁴Rosetta Science Operations Centre, European Space Astronomy Centre, 28691 Villanueva de la Cañada, Madrid, Spain. ⁵Astronomy Department, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁶Space Research Centre, Polish Academy of Sciences, 00-716 Warsaw, Poland.

Table 1 | Calculating the D/H ratio in water in comet 103P/Hartley 2

Model	T_{gas} (K)	x_{ne}	$\langle N(\text{HDO}) \rangle$ (10^{10} cm^{-2})	$Q(\text{HDO})$ (10^{24} s^{-1})	$\langle N(\text{H}_2^{18}\text{O}) \rangle$ (10^{11} cm^{-2})	$Q(\text{H}_2^{18}\text{O})$ (10^{25} s^{-1})	D/H
(1)	50	0.	4.9	3.1	3.5	2.1	1.49×10^{-4}
	50	0.2	3.6	2.3	2.5	1.5	1.55×10^{-4}
	70	0.2	3.7	2.4	2.5	1.5	1.60×10^{-4}
	Law	0.1	5.7	3.6	3.7	2.2	1.63×10^{-4}
(2)	50	0.2	4.3	2.7	2.9	1.8	1.54×10^{-4}
	Law	0.1	4.8	3.1	3.2	1.9	1.58×10^{-4}

The parameter x_{ne} , scaling the electron density profile in the models, is constrained by mapping observations. $\langle N \rangle$ and Q are respectively beam integrated column density and production rate, determined using different parameters in the excitation models^{14,29}. Production rates were computed assuming isotropic outflow of water from the nucleus, with a velocity of 0.6 km s^{-1} , consistent with the width of the H_2^{18}O line. We accounted for the $10''$ offset between the centre of the beam and the position of the peak of the H_2O distribution. Values of 50 and 70 K for the gas kinetic temperature, T_{gas} , are consistent with multi-transition measurements of gaseous species in the millimetre and near-infrared range, respectively. The gas kinetic temperature is expected to decrease with increasing distance owing to quasi-adiabatic expansion of the escaping gases: the temperature law assumes that $T_{\text{gas}} = 80 \text{ K}$ for $r < 270 \text{ km}$, $T_{\text{gas}} = 12 \text{ K}$ for $r > 630 \text{ km}$, with a linear decrease between 270 and 630 km, where r is the distance from the nucleus. Collision cross-sections involving water molecules and electrons are modelled differently in models (1) and (2). Both models use an electron density profile based on *in situ* measurements in comet 1P/Halley scaled to the activity of 103P/Hartley 2 (ref. 29). The D/H ratio is equal to $0.5 \times Q(\text{HDO})/Q(\text{H}_2\text{O})$, with $Q(\text{H}_2\text{O}) = 500 \times Q(\text{H}_2^{18}\text{O})$. See Supplementary information for details of the models and model parameters.

comparable to the primordial D/H ratio in the Universe after the Big Bang (Fig. 2)¹⁵. Protosolar water, on the other hand, is believed to be highly enriched (D/H $\approx 1 \times 10^{-3}$)¹⁶ due to the low-temperature (~ 10 – 30 K) non-equilibrium chemistry that characterizes the dense interstellar medium¹⁷, either via gas-phase isotopic exchange reactions involving ions and radicals, or grain-surface processes. Consequently, the resulting D/H ratio in water ice is very sensitive to the physical conditions, in particular the kinetic temperature of the medium. After the protosolar cloud collapsed to form the solar nebula, isotopic exchange reactions between molecular hydrogen and HDO molecules would have led to a gradual reduction of D/H in water¹⁸, as compared to the initial interstellar medium value. Because the efficiency of these reactions and the turbulent mixing within the solar nebula is correlated with the gas density and temperature, the deuterium enhancement in water has been predicted to increase with the heliocentric distance^{19–21}. Ices, captured by planetesimals and cometesimals, would have then preserved the deuterium enrichment in water from this early epoch. As

a result, small Solar System bodies are expected to exhibit different D/H ratios in their water ice depending on the distance from the Sun at which they were formed.

In the context of this simple nebular model, the D/H ratio of $(1.61 \pm 0.24) \times 10^{-4}$ in comet 103P/Hartley 2—a factor of two lower than that measured in Oort-cloud comets (Fig. 2) and, within uncertainties, consistent with that of the Earth's oceans (for which the Vienna Standard Mean Ocean Water (VSMOW) value is $(1.558 \pm 0.001) \times 10^{-4}$)—is therefore surprising, and compatible with two different schemes: (1) either this comet did not form in a region that was further from the Sun than the assembly zone of the Oort-cloud comets, or (2) the dependence of the water D/H ratio with distance from the Sun is not as expected on the basis of current models. Concerning the first possibility, dynamical models indeed suggest that a fraction of the Jupiter-family comets originate in the Oort cloud²². Still, even if comet 103P/Hartley 2 stems from the Oort cloud, this would not explain why its D/H ratio is different from that seen in other Oort-cloud comets. Models also suggest that a fraction of the Jupiter-family comets may have originated from the Trojan asteroid swarms sharing the orbit of Jupiter²³. The Trojans are generally thought to have resided at their current location since the formation of the Solar System. Therefore, Jupiter family comets originating in the Trojan region could, in theory, display deuterium enrichment values lower than those for bodies originating in the Kuiper belt, if they indeed formed in the vicinity of Jupiter. However, the most probable scenario is that 103P/Hartley 2 originated in the Kuiper belt.

It is difficult to explain the low D/H ratio in 103P/Hartley 2 (compared to that of previous measurements in comets) with the formation regions of comets, thus models of the gradient of D/H in the Solar System—predictions not yet directly confirmed by observations, owing to scarcity of accurate isotopic measurements—may need to be revisited. In fact, one recent model has suggested that the D/H ratio of water vapour can be locally enhanced²⁴. However, the vapour must then be implanted into cometary ices. Moreover, until the measurement of 103P/Hartley 2 there was no observational confirmation of variations in the D/H ratio. One possible solution is that there was large-scale movement of material between the inner and outer Solar

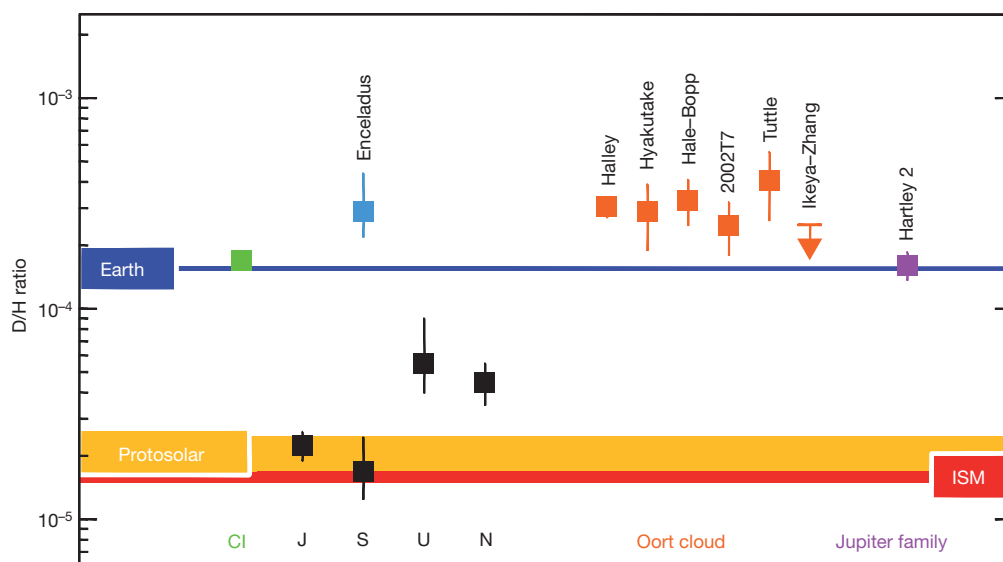


Figure 2 | D/H ratios in the Solar System. Orange squares, values measured for water in the Oort-cloud comets 1P/Halley, C/1996 B2 (Hyakutake), C/1995 O1 (Hale-Bopp), C/2002 T7 (LINEAR) and 8P/Tuttle. Arrow (for 153P/Ikeya-Zhang), upper limit. Purple square, present measurement in the water of 103P/Hartley 2. Black symbols, D/H ratio in H_2 in the atmosphere of the giant planets—Jupiter (J), Saturn (S), Uranus (U) and Neptune (N). Light blue and green symbols, D/H values for water in the plume of Saturn's moon Enceladus and in CI carbonaceous chondrites, respectively. Error bars, 1σ . The D/H

determinations in comets originating from the Oort cloud are twice the value for the Earth's ocean (blue line) and about a factor of ten larger than the protosolar value in H_2 (broad yellow line), the latter being comparable to the value in atomic hydrogen found in the local interstellar medium (ISM, red horizontal line). The D/H ratio in the Jupiter-family comet 103P/Hartley 2 is the same as the Earth's ocean value and the chondritic CI value. Uranus and Neptune have been enriched in deuterium by the mixing of their atmospheres with D-rich protoplanetary ices. For further details, see Supplementary Table 1.

System. According to a recent theory proposed for the early Solar System (the Grand Tack scenario), when the giant planets were still embedded in the nebular gas disk, there was a general radial mixing of the distribution of comets and asteroids born in different regions²⁵. The similarity of the D/H ratio in comet 103P/Hartley 2, which probes the Kuiper belt, with that found in CI chondrites tracing the asteroid belt, would be in agreement with a general shake-up of the Solar System at early times.

In more 'orderly' models, the high D/H values derived from the earlier observations of Oort-cloud comets suggested that at most 10% of the Earth's water could have been supplied from the outermost Solar System, but even under these circumstances a number of scenarios have been developed suggesting that terrestrial water could have in fact been delivered by comets. Such models are based on assumptions about the heliocentric D/H gradient²⁶, and the analysis of lunar samples²⁷ and telluric sedimentary rocks formed at the end of the Late Heavy Bombardment phase²⁸. Our Herschel observations of a VSMOW-like D/H ratio in 103P/Hartley 2 enlarge the region of the solar nebula known to have a D/H ratio similar to that of Earth's oceans; this region now includes both the asteroid belt and the much larger Kuiper belt, thereby providing support for the theory of a common water source for the inner Solar System bodies (including the Earth) in which comets play an important part.

Further constraints on the delivery of volatiles to the early Earth and an improved understanding of the origin of the different dynamical classes of comets will require significantly larger sample sizes than those at present available. A handful of additional measurements can be expected from Herschel before its cryogen supply is exhausted, but the comparison of D/H ratios in the inner and outer Solar System must necessarily utilize very different objects and materials. For the inner Solar System, *in situ* space missions or sample return missions to the outer asteroid belt would provide critical new data. Astronomically, the minuscule strength of HDO spectroscopic signatures makes D/H measurements extremely challenging, and dedicated programmes using new facilities will be required to substantially increase the inventory of high-precision D/H ratios in comets and other icy Solar System bodies, including the Jovian satellites.

Received 16 May; accepted 30 August 2011.

Published online 5 October 2011.

- Javoy, M. *et al.* The chemical composition of the Earth: enstatite chondrite models. *Earth Planet. Sci. Lett.* **293**, 259–268 (2010).
- Wänke, H. Constitution of terrestrial planets. *Phil. Trans. R. Soc. Lond.* **303**, 287–302 (1981).
- Robert, F. A distinct source of lunar water? *Nature Geosci.* **4**, 74–75 (2011).
- Drake, M. J. & Righter, K. Determining the composition of the Earth. *Nature* **416**, 39–44 (2002).
- Oró, J. Comets and the formation of biochemical compounds on the primitive Earth. *Nature* **190**, 389–390 (1961).
- Owen, T., Bar-Nun, A. & Kleinfeld, I. Possible cometary origin of heavy noble gases in the atmospheres of Venus, Earth and Mars. *Nature* **358**, 43–46 (1992).
- Morbidelli, A. *et al.* Source regions and time scales for the delivery of water to Earth. *Meteorit. Planet. Sci.* **35**, 1309–1320 (2000).
- Tsiganis, K., Gomes, R., Morbidelli, A. & Levison, H. F. Origin of the orbital architecture of the giant planets of the Solar System. *Nature* **435**, 459–461 (2005).
- Gomes, R., Levison, H. F., Tsiganis, K. & Morbidelli, A. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets. *Nature* **435**, 466–469 (2005).
- Levison, H. F. & Duncan, M. J. From the Kuiper belt to Jupiter-family comets: the spatial distribution of ecliptic comets. *Icarus* **127**, 13–32 (1997).
- Dones, L., Weissman, P. R., Levison, H. F. & Duncan, M. J. in *Comets II* (eds Festou, M. C., Keller, H. U. & Weaver, H. A.) 153–174 (Univ. Arizona Press, 2005).
- Hartogh, P. *et al.* Water and related chemistry in the solar system: a guaranteed time key programme for Herschel. *Planet. Space Sci.* **57**, 1596–1606 (2009).

- Meech, K. *et al.* EPOXI: 103P/Hartley 2 observations from a worldwide campaign. *Astrophys. J.* **734**, L1 (2011).
- Biver, N. *et al.* Submillimetre observations of comets with Odin: 2001–2005. *Planet. Space Sci.* **55**, 1058–1068 (2007).
- Linsky, J. L. *et al.* What is the total deuterium abundance in the local galactic disk? *Astrophys. J.* **647**, 1106–1124 (2006).
- Butner, H. M. *et al.* Discovery of interstellar heavy water. *Astrophys. J.* **659**, L137–L140 (2007).
- Watson, W. D. Ion-molecule reactions, molecule formation, and hydrogen-isotope exchange in dense interstellar clouds. *Astrophys. J.* **188**, 35–42 (1974).
- Geiss, J. & Reeves, H. Cosmic and solar system abundances of deuterium and helium-3. *Astron. Astrophys.* **18**, 126–132 (1972).
- Drouart, A., Dubrulle, B., Gautier, D. & Robert, F. Structure and transport in the solar nebula from constraints on deuterium enrichment and giant planet formation. *Icarus* **140**, 129–155 (1999).
- Mousis, O. *et al.* Constraints on the formation of comets from D/H ratios measured in H₂O and HCN. *Icarus* **148**, 513–525 (2000).
- Aikawa, Y. & Herbst, E. Two-dimensional distributions and column densities of gaseous models in protoplanetary disks II. Deuterated species and UV shielding by ambient clouds. *Astron. Astrophys.* **371**, 1107–1117 (2001).
- Emel'yanenko, V. V., Asher, D. J. & Bailey, M. E. Centaurs from the Oort cloud and the origin of Jupiter-family comets. *Mon. Not. R. Astron. Soc.* **361**, 1345–1351 (2005).
- Marzari, F., Farinella, P. & Vanzani, V. Are Trojan collisional families a source for short-period comets? *Astron. Astrophys.* **299**, 267–276 (1995).
- Thi, W.-F., Woitke, P. & Kamp, I. Warm non-equilibrium gas phase chemistry as a possible origin of high HDO/H₂O ratios in hot and dense gases: application to inner protoplanetary discs. *Mon. Not. R. Astron. Soc.* **407**, 232–246 (2010).
- Walsh, K. J. & Morbidelli, A. The effect of an early planetesimal-driven migration of the giant planets on terrestrial planet formation. *Astron. Astrophys.* **526**, A126 (2011).
- Delsemme, A. H. The deuterium enrichment observed in recent comets is consistent with the cometary origin of seawater. *Planet. Space Sci.* **47**, 125–131 (1998).
- Greenwood, J. P. *et al.* Hydrogen isotope ratios in lunar rocks indicate delivery of cometary water to the Moon. *Nature Geosci.* **4**, 79–82 (2011).
- Jørgensen, U. G. *et al.* The Earth-Moon system during the late heavy bombardment period — geochemical support for impacts dominated by comets. *Icarus* **204**, 368–380 (2009).
- Zakharov, V., Bockelée-Morvan, D., Biver, N., Crovisier, J. & Lecacheux, A. Radiative transfer simulation of water rotational excitation in comets. Comparison of the Monte Carlo and escape probability methods. *Astron. Astrophys.* **473**, 303–310 (2007).
- Bensch, F. & Bergin, E. A. The pure rotational line emission of ortho-water in comets. I. Radiative transfer model. *Astrophys. J.* **615**, 531–544 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Herschel is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation by NASA. The Heterodyne Instrument for the Far Infrared (HIFI) has been designed and built by a consortium of institutes and university departments from across Europe, Canada and the United States under the leadership of SRON, the Netherlands Institute for Space Research, and with major contributions from Germany, France and the USA. This development has been supported by national funding agencies: CEA, CNES, CNRS (France); ASI (Italy); and DLR (Germany). Additional funding support for some instrument activities has been provided by ESA. Support for this work was also provided by NASA through an award issued by JPL/Caltech. D.C.L. is supported by an NSF award to the Caltech Submillimeter Observatory. We thank R. Lorente, P. García-Lario, M. Kidger and G. Pilbratt for helping with the scheduling of these observations, and I. Avruch for the assistance with HIFI specific software issues.

Author Contributions This paper represents the combined work of the HsO (the Herschel guaranteed time key programme "Water and related chemistry in the solar system") team members listed as authors. P.H. is the coordinator of this programme. All authors contributed to this work, including observation planning, data analysis and writing of the manuscript. N.B., D.B.-M., M.R., R.M., M.d.V.-B. and M.E. carried out the data reduction and contributed to the modelling efforts. All authors were collectively involved in the discussion and interpretation of the results.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.H. (hartogh@mps.mpg.de).

Unprecedented Arctic ozone loss in 2011

Gloria L. Manney^{1,2}, Michelle L. Santee¹, Markus Rex³, Nathaniel J. Livesey¹, Michael C. Pitts⁴, Pepijn Veefkind^{5,6}, Eric R. Nash⁷, Ingo Wohltmann³, Ralph Lehmann³, Lucien Froidevaux¹, Lamont R. Poole⁸, Mark R. Schoeberl⁹, David P. Haffner⁷, Jonathan Davies¹⁰, Valery Dorokhov¹¹, Hartwig Gernandt³, Bryan Johnson¹², Rigel Kivi¹³, Esko Kyrö¹³, Niels Larsen¹⁴, Pieter F. Levelt^{5,6,15}, Alexander Makshtas¹⁶, C. Thomas McElroy¹⁰, Hideaki Nakajima¹⁷, Maria Concepción Parrondo¹⁸, David W. Tarasick¹⁰, Peter von der Gathen³, Kaley A. Walker¹⁹ & Nikita S. Zinoviev¹⁶

Chemical ozone destruction occurs over both polar regions in local winter–spring. In the Antarctic, essentially complete removal of lower-stratospheric ozone currently results in an ozone hole every year, whereas in the Arctic, ozone loss is highly variable and has until now been much more limited. Here we demonstrate that chemical ozone destruction over the Arctic in early 2011 was—for the first time in the observational record—comparable to that in the Antarctic ozone hole. Unusually long-lasting cold conditions in the Arctic lower stratosphere led to persistent enhancement in ozone-destroying forms of chlorine and to unprecedented ozone loss, which exceeded 80 per cent over 18–20 kilometres altitude. Our results show that Arctic ozone holes are possible even with temperatures much milder than those in the Antarctic. We cannot at present predict when such severe Arctic ozone depletion may be matched or exceeded.

Since the emergence of the Antarctic ‘ozone hole’ in the 1980s¹ and elucidation of the chemical mechanisms^{2–5} and meteorological conditions⁶ involved in its formation, the likelihood of extreme ozone depletion over the Arctic has been debated. Similar processes are at work in the polar lower stratosphere in both hemispheres, but differences in the evolution of the winter polar vortex and associated polar temperatures have in the past led to vastly disparate degrees of spring-time ozone destruction in the Arctic and Antarctic. We show that chemical ozone loss in spring 2011 far exceeded any previously observed over the Arctic. For the first time, sufficient loss occurred to reasonably be described as an Arctic ozone hole.

Arctic polar processing in 2010–11

In the winter polar lower stratosphere, low temperatures induce condensation of water vapour and nitric acid (HNO₃) into polar stratospheric clouds (PSCs). PSCs and other cold aerosols provide surfaces for heterogeneous conversion of chlorine from longer-lived reservoir species, such as chlorine nitrate (ClONO₂) and hydrogen chloride (HCl), into reactive (ozone-destroying) forms, with chlorine monoxide (ClO) predominant in daylight^{5,7}.

In the Antarctic, enhanced ClO is usually present for 4–5 months (through to the end of September)^{8–11}, leading to destruction of most of the ozone in the polar vortex between ~14 and 20 km altitude⁷. Although ClO enhancement comparable to that in the Antarctic occurs at some times and altitudes in most Arctic winters⁹, it rarely persists for more than 2–3 months, even in the coldest years¹⁰. Thus chemical ozone loss in the Arctic has until now been limited, with largest previous losses observed in 2005, 2000 and 1996^{7,12–14}.

The 2010–11 Arctic winter–spring was characterized by an anomalously strong stratospheric polar vortex and an atypically long continuously cold period. In February–March 2011, the barrier to

transport at the Arctic vortex edge was the strongest in either hemisphere in the last ~30 years (Fig. 1a, Supplementary Discussion).

The persistence of a strong, cold vortex from December through to the end of March was unprecedented. In the previous years with most ozone loss, temperatures (T) rose above the threshold associated with chlorine activation (T_{act} , near 196 K, roughly the threshold for the potential existence of PSCs) by early March (Fig. 1b, Supplementary Figs 1, 2). Only in 2011 and 1997 have Arctic temperatures below T_{act} persisted through to the end of March, sporadically approaching a vortex volume fraction similar in size to that in some Antarctic winters (Fig. 1b). In 1996–97, however, the cold volume remained very limited until mid-January and was smaller than that in 2011 at most times during late January through to the end of March (Fig. 1b, Supplementary Figs 1, 2).

Daily minimum temperatures in the 2010–11 Arctic winter were not unusually low, but the persistently cold region was remarkably deep (Supplementary Figs 1, 2). Temperatures were below T_{act} for more than 100 days over an altitude range of ~15–23 km, compared to a similarly prolonged cold period over only ~20–23 km altitude in 1997; below ~19 km altitude, $T < T_{\text{act}}$ continued for ~30 days longer in 2011 than in 1997 (Supplementary Fig. 1b). In 2005, the previous year with largest Arctic ozone loss⁷, $T < T_{\text{act}}$ occurred for more than 100 days over ~17–23 km altitude, but all before early March.

The winter mean volume of air in which PSCs may form (that is, with $T < T_{\text{act}}$), V_{psc} , is closely correlated with the potential for ozone loss^{7,15–17}. In 2011, V_{psc} (as a fraction of the vortex volume) was the largest on record (Fig. 1c). Both large V_{psc} and cold lingering well into spring are important in producing severe chemical loss^{7,15,16}, and 2010–11 was the only Arctic winter during which both conditions have been met. Much lower fractional V_{psc} in 1997 than in 1996, 2000, 2005 or 2011 (Fig. 1c) is consistent with less ozone loss that year^{16,17}.

¹Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. ²New Mexico Institute of Mining and Technology, Socorro, New Mexico 87801, USA. ³Alfred Wegener Institute for Polar and Marine Research, D-14473 Potsdam, Germany. ⁴NASA Langley Research Center, Hampton, Virginia 23681, USA. ⁵Royal Netherlands Meteorological Institute, 3730 AE De Bilt, The Netherlands. ⁶Delft University of Technology, 2600 GA Delft, The Netherlands. ⁷Science Systems and Applications, Inc., Lanham, Maryland 20706, USA. ⁸Science Systems and Applications, Inc., Hampton, Virginia 23666, USA. ⁹Science and Technology Corporation, Lanham, Maryland 20706, USA. ¹⁰Environment Canada, Toronto, Ontario, Canada M3H 5T4. ¹¹Central Aerological Observatory, Dolgoprudny 141700, Russia. ¹²NOAA Earth System Research Laboratory, Boulder, Colorado 80305, USA. ¹³Arctic Research Center, Finnish Meteorological Institute, 99600 Sodankylä, Finland. ¹⁴Danish Climate Center, Danish Meteorological Institute, DK-2100 Copenhagen, Denmark. ¹⁵Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands. ¹⁶Arctic and Antarctic Research Institute, St Petersburg 199397, Russia. ¹⁷National Institute for Environmental Studies, Tsukuba-city, 305-8506, Japan. ¹⁸National Institute for Aerospace Technology, 28850 Torrejón De Ardoz, Spain. ¹⁹University of Toronto, Toronto, Ontario, Canada M5S 1A7.

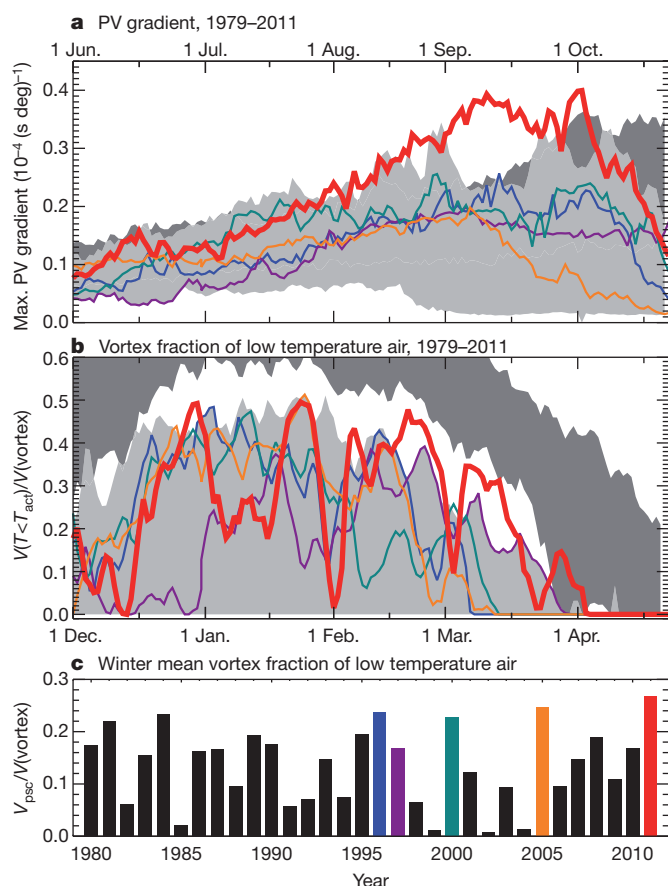


Figure 1 | Meteorology of the Arctic lower stratosphere. **a**, Vortex strength (as indicated by maximum potential vorticity⁴⁹ (PV) gradients) at 460 K potential temperature (~ 18 km altitude, ~ 65 hPa level). **b**, Fraction of vortex volume at potential temperatures between 390 and 550 K with a temperature less than the chlorine activation threshold (T_{act}). Light (dark) grey shading shows range of Arctic (Antarctic) values for 1979–2010. Antarctic dates are shifted by six months (top axis in **a**) to show the equivalent season. **c**, Winter mean V_{psc} during the past 32 years, expressed as a fraction of vortex volume. Red, orange, green, purple and blue lines/bars show the 2010–11, 2004–05, 1999–2000, 1996–97 and 1995–96 Arctic winters, respectively.

Factors playing secondary parts in governing interannual variability in ozone destruction, including vortex strength, structure and position relative to the cold region, also favour large loss in 2011 (Supplementary Figs 2, 3, Supplementary Discussion). However, despite the fraction of the vortex with $T < T_{act}$ and mid-March temperatures sporadically approaching those seen in the Antarctic (Fig. 1b, Supplementary Fig. 1a), even in 2011 temperatures were much higher, and the cold regions much smaller, than those in most Antarctic winters.

Satellite trace-gas and PSC measurements highlight the stark contrast between polar processing in 2010–11 and that in typical Arctic winters, and the parallels with Antarctic conditions (Figs 2, 3). In 2011, PSCs or aerosols were abundant until mid-March (Fig. 3a; consistent with a deep region with $T < T_{act}$, Fig. 3b), much later than usual in the Arctic^{18–20}, with vortex-average amounts at some altitudes similar to those in the Antarctic and dramatically larger than the near-zero values at that time in most Arctic winters. Furthermore, PSCs in 2011 spanned an altitude range comparable to that in the Antarctic, an uncommon occurrence in the Arctic^{18–20}. Particles in long-lasting PSCs can grow large enough to sediment, resulting in denitrification, permanent removal of HNO_3 from the stratosphere^{7,12}. By late March 2011 no PSCs remained (Fig. 3a), yet HNO_3 mixing ratios were much lower than observed in any previous Arctic winter (Fig. 2a). The continuing depression in HNO_3 after PSCs had evaporated indicates

denitrification. Albeit less severe than in typical Antarctic winters (Fig. 2b, c, 3c), the extent and degree of denitrification in 2011 were unmatched in the Arctic, approaching the range of Antarctic conditions for the first time.

Decreasing HCl and increasing ClO signify chlorine activation (Fig. 2d–i). Some ClO enhancement has occurred in all recent Arctic winters, but has never been as prolonged and extensive as that in 2011. In late February, high ClO pervaded the sunlit portion of the vortex. The 2011 values vastly exceed the range previously observed in the Arctic from late February through to the end of March. They also briefly lie outside the Antarctic seasonal envelope, primarily because the higher solar zenith angles of the Antarctic measurements used here lead to $\sim 30\%$ lower ClO under fully activated conditions. In late February, HCl values (unaffected by solar zenith angle issues) fall along the lower boundary of the Antarctic envelope, confirming the picture seen in ClO. The vertical extent of chlorine activation was also comparable to that in the Antarctic (Fig. 3d, e).

In previous cold Arctic winters, chlorine was deactivated (converted from ozone-destroying forms into less reactive reservoir species) by mid-March¹¹; even in 1997, ClO started to decline by late February (Fig. 2g). In 2011, by contrast, ClO began decreasing rapidly only about a week earlier than is typical in the Antarctic. ClO data in late February 1997 indicate that not only were maximum values lower than those in early March 2011, but also the vertical range of enhancement was shallower, with weaker activation at low altitudes than in 2011 (Fig. 3e), consistent with the higher altitudes and decreasing extent (Figs 1b, 3b, Supplementary Fig. 2) of $T < T_{act}$.

When chlorine is deactivated, whether it is converted first into HCl or ClONO_2 depends sensitively upon HNO_3 and ozone abundances. In the Arctic, chlorine is normally deactivated through initial reformation of ClONO_2 . In the severely denitrified and ozone-depleted Antarctic vortex, production of ClONO_2 is suppressed and that of HCl highly favoured^{11,12,21}. In March 2011, the recovery of HCl followed a much more Antarctic-like pathway than has been observed in any other Arctic winter.

The largest Arctic chemical ozone loss was previously observed in 2005, followed closely by 2000 and 1996^{7,12–14}. Although low temperatures persisted until the end of March 1997, the ozone loss in that year was far less. No previous year rivals 2011, when the evolution of Arctic ozone more closely followed that typical of the Antarctic (Fig. 2j). Ozone profiles in late March 2011 resemble typical Antarctic late-winter profiles much more strongly than they do the average Arctic one (Fig. 3f). Because mixing in April 2011 (for example, lamination events larger than that shown in Fig. 3f) entrained ozone-rich air into the vortex, the slight decrease in vortex-averaged ozone at a potential temperature of 485 K from 26 March to 20 April (from ~ 1.8 to ~ 1.6 p.p.m.v., Fig. 2j) indicates continuing chemical loss during this interval.

Estimates of chemical ozone loss

Chemical loss is difficult to quantify in the Arctic, where transport from above replenishes ozone in the lower stratospheric vortex, obscuring the signature of chlorine-catalysed destruction^{12,22,23}. The evolution of the long-lived trace gas nitrous oxide (N_2O) reflects steady downward transport throughout the 2010–11 winter–spring, indicating that subsidence partially masked chemical loss. Horizontal transport can also confound the signature of chemical loss, bringing air into the vortex that has either higher²⁴ or lower¹⁴ concentrations of ozone, depending on the altitude and latitude from which it originates.

Representative results from two types of chemical loss calculations^{24–28} based on balloon-borne and satellite observations are shown in Fig. 4. The differences (up to ~ 0.4 p.p.m.v. at the end of March 2011) in estimates derived from the various methods and data sets imply some uncertainty in the chemical loss determination. Year-to-year differences in the amount of ozone loss are very similar when obtained from any method/data set combination, however, indicating a high degree of precision in the relative amount of calculated loss

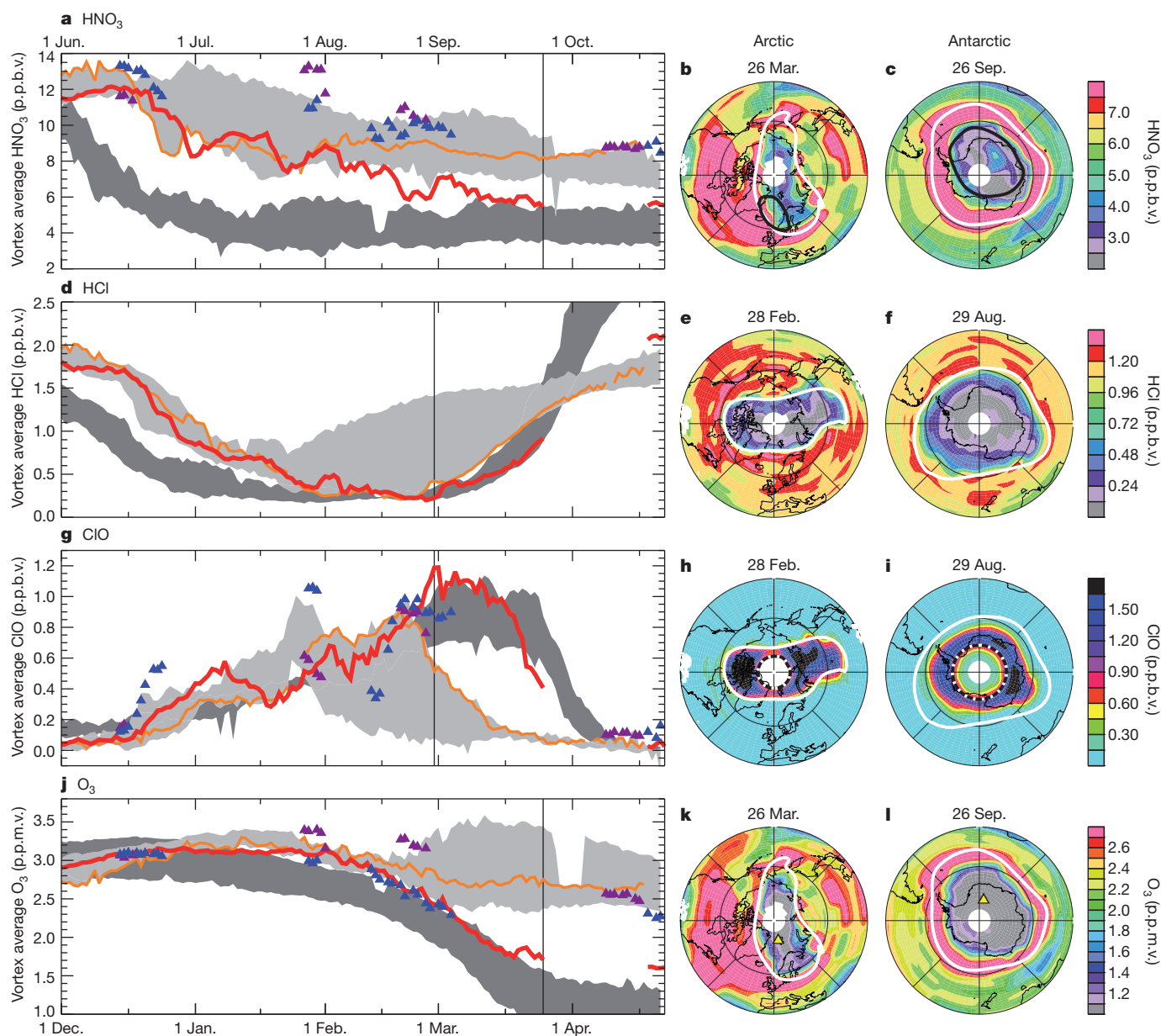


Figure 2 | Chemical composition in the lower stratosphere. a–l, Maps (right) and vortex-averaged time series (left) at 485 K potential temperature (~ 20 km, ~ 50 hPa) for four different gases: HNO_3 (a, b, c), HCl (d, e, f), ClO (g, h, i) and O_3 (ozone; j, k, l); mixing ratios from Aura MLS are shown. Averaging for the time series is done within the white contour shown on the maps. Blue (purple) triangles on time series, 1995–96 (1996–97) values from UARS MLS. Line colours/shading as in Fig. 1, but shading is for Aura MLS measurements from

2005–10. Antarctic dates are shifted by six months (top axis on time series) to show the equivalent season. Vertical lines show dates of maps in 2011 (2010) in the Arctic (Antarctic). Black overlays on HNO_3 maps, T_{act} (~ 196 K at this level); HNO_3 may be sequestered in PSCs at lower temperatures. Dotted black/white contour on ClO maps, 92° SZA, poleward of which measurements were taken in darkness. Yellow/black triangles on ozone maps, locations of the profiles in Fig. 3.

between different years. Chemical destruction was severe between ~ 16 and 22 km altitude, with the largest loss exceeding 2.5 p.p.m.v. by 26 March 2011 (Fig. 4a). By 31 March 2011, chemical loss was nearly double that in 2005 from ~ 18 to above 22 km, and similar to that in 2005 at lower altitudes (Fig. 4b, c). From ~ 18 to 20 km, more than 80% of the ozone present in January had been chemically destroyed by late March. Chemical removal in 1996 and 2000 started at a rate similar to that in 2011 (Fig. 4c), but ceased by late March; maximum losses in 2000 approached those in 2011, but extended over a much smaller vertical range (Fig. 4b). Loss in 1996, 2000 and 2005 considerably exceeded that in 1997, with greater destruction at lower altitudes in those years contributing more to total column loss^{7,12,13}. Chemical loss in 2011 was two to three times larger than that in 1997, and about twice that in 1996 and 2005 above ~ 16 km; from ~ 15 to 23 km it was comparable to that in the Antarctic ozone hole in 1985²⁹.

Single ozone-sonde station measurements in early April 2011 suggest continuing ozone loss (Fig. 4c).

Although the meteorology during March–April was similar in 1997 and 2011, ozone loss was much more pronounced in 2011. Photochemical box model simulations (Supplementary Fig. 4, Supplementary Discussion) elucidate how early winter conditions set the stage for record springtime ozone destruction in 2011. Chlorine activation brought on by enduring cold from December through to the end of February led to ~ 0.7 – 0.8 p.p.m.v. lower ozone at the beginning of March 2011 (Figs 2j, 4c). The early onset of continuous cold also facilitated formation of PSC particles large enough to sediment, resulting in ~ 4 p.p.b.v. less HNO_3 by March in 2011 than in 1997 (Fig. 2a). The degree of denitrification has a profound impact on the severity of springtime Arctic ozone loss³⁰. By delaying chlorine deactivation, lower HNO_3 by 1 March was responsible for ~ 0.6 p.p.m.v. more ozone

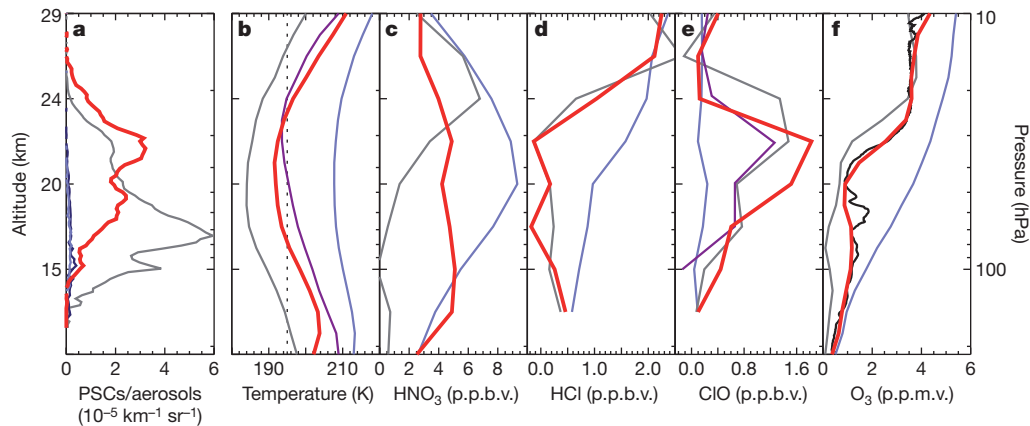


Figure 3 | Vertical composition information. **a**, Red, PSCs/aerosol amounts averaged in the vortex over a week centred around 25 February 2011; dark blue, the average for the same week in 2007–10; grey, the average over the equivalent period (centred on 28 August) for the Antarctic in 2006–10; lavender, the Arctic average for a week centred around 26 March 2011. (In late winter–spring, maximum PSC altitudes are generally higher in the Arctic because early winter PSC activity redistributes HNO_3 and water vapour to lower altitudes in the Antarctic¹⁸). **b–f**, Daily average profiles of MERRA temperatures (**b**) and MLS HNO_3 (**c**), HCl (**d**), ClO (**e**) and ozone (**f**). Red lines, data from a $4^\circ \times 15^\circ$ latitude \times longitude box around 79°N , 12°E ; in **c**, **f**, taken on 26 March; in

b, **d**, **e**, on 6 March 2011. Lavender, 7-day average for 2005–10 (1980–2010 for **b**) centred on the same location and days. Grey, profiles in a similar box in the Antarctic (79°S , 12°E) on 26 September for **c**, **f**, and on 8 September 2010 for **b**, **d**, **e**. Dotted black line in **b**, approximate T_{act} (195 K), see text. Purple line in **b**, 7-day average around 6 March 1997, centred on same location. Purple line in **e**, a midday ClO profile from UARS MLS on 26 February 1997 averaged in an $8^\circ \times 30^\circ$ box centred at the same Arctic location. A high-resolution ozone-sonde profile at Ny Ålesund on 26 March 2011 (black in **f**) agrees well with MLS; lamination, a signature of mixing with ozone-rich extra-vortex air, is apparent as a local maximum near 60 hPa.

loss after that date in 2011 than in 1997 (Supplementary Fig. 4, Supplementary Discussion). The effects of denitrification and early-winter loss together account for the disparity in ozone depletion in these two winters (~ 1.5 p.p.m.v. more loss at 460 K in 2011 than in 1997, Fig. 4c, Supplementary Fig. 4). Loss as severe as that in 2011 thus

requires $T < T_{\text{act}}$ with consequent chlorine activation and ozone destruction, early in winter (as in 1996, 2000 and 2005, but not in 1997), a cold period and region before March sufficient to allow widespread denitrification, and the persistence of a cold polar vortex into April (as in 1997, but not in 1996, 2000 or 2005).

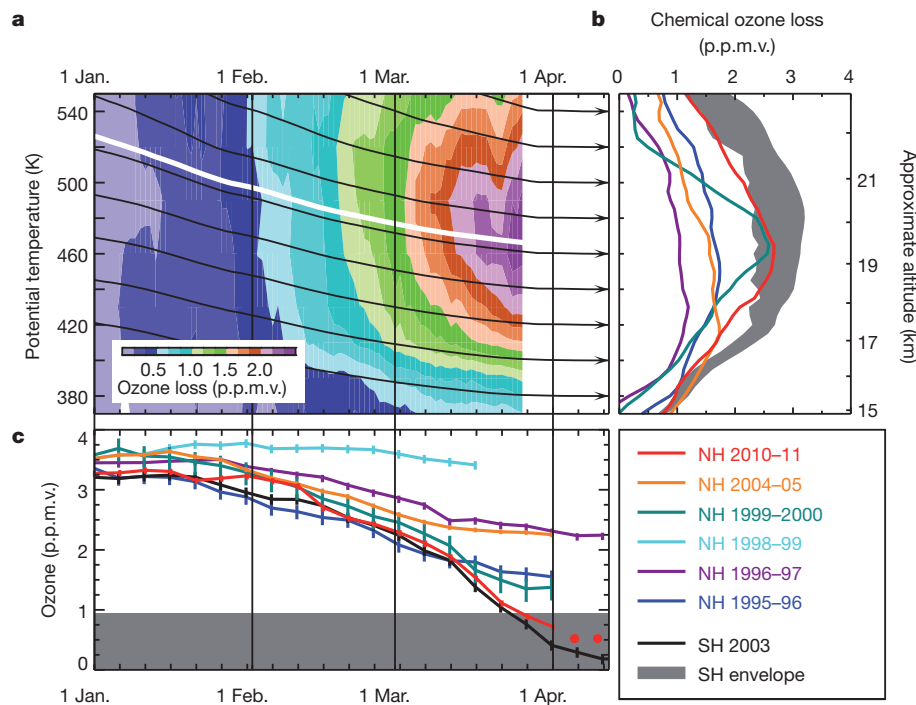


Figure 4 | Chemical ozone loss estimates. **a**, Chemical loss as a function of time and potential temperature from passive subtraction of MLS and ATLAS passively-transported ozone (initialized with December MLS data). **b**, Chemical loss from ozone sondes in unmixed vortex air as a function of 'spring equivalent potential temperature'⁴⁸ (black contours in **a**). Shading, Antarctic range defined by 1985 (the first year with profile measurements inside the ozone hole²⁹) and 2003 (a recent year with a severe ozone hole). The 2003 Antarctic curve is shifted by six months minus 10 days because ozone sondes that year predominantly sampled the outermost vortex, where ozone

loss begins earliest. **c**, Ozone at a spring equivalent potential temperature of 465 K (white contour in **a**), near the level of maximum chemical loss. Shading, the region below the minimum reached in the 1985 Antarctic ozone hole. In April 2011 most soundings sampled the disturbed vortex edge; only two were made in air uninfluenced by mixing (red dots). Error bars, 1σ uncertainties based on the scatter of individual ozone-sonde measurements. Line colours as in Fig. 1; 1998–99 (a winter with no ozone loss) is shown in cyan. NH, Northern Hemisphere; SH, Southern Hemisphere.

Column ozone

Total column ozone is a predominant factor determining exposure of Earth's surface to ultraviolet radiation^{7,12}. In the context of previous Arctic winters, 2011 was truly remarkable: the fraction of the Arctic vortex in March with total ozone less than 275 Dobson units (DU) is typically near zero, but reached nearly 45% in 2011 (Fig. 5a). Because of the dynamically-driven correlation between total ozone and lower-stratospheric temperature^{23,31–34} (Supplementary Discussion), the abiding cold in 1997 and 2011 would have led to lower March total ozone than in other Arctic winters even without chemical loss; dynamical conditions in March–April 1997 particularly favoured low total ozone³³ (Supplementary Discussion). In March 2011, however, the area of low total ozone covered more than twice as much of the vortex as in 1997, and the daily vortex 'ozone deficit' (Supplementary Fig. 5a) was 30–50 DU larger, consistent with the greater chemical loss (Fig. 4). Maximum 2011 vortex fractions of low ozone approached those in early Antarctic ozone holes (Fig. 5a). The close correspondence between the vortex and both low total ozone and the large Arctic total ozone deficit (Fig. 5b, d) implies that low total ozone in March 2011 resulted primarily from chemical loss^{31,32} (Supplementary Discussion). The ozone deficit in the Antarctic (Fig. 5e) shows a maximum over 0–90° W, and a minimum over 90–200° E, reflecting a vortex position in 2010 different to that in the reference state (which is less robust than that for the Arctic). Differences in morphology deep in the vortex are, however, minimal. The 2011 Arctic ozone deficit was at least comparable to that in the 2010 Antarctic vortex core at an equivalent time.

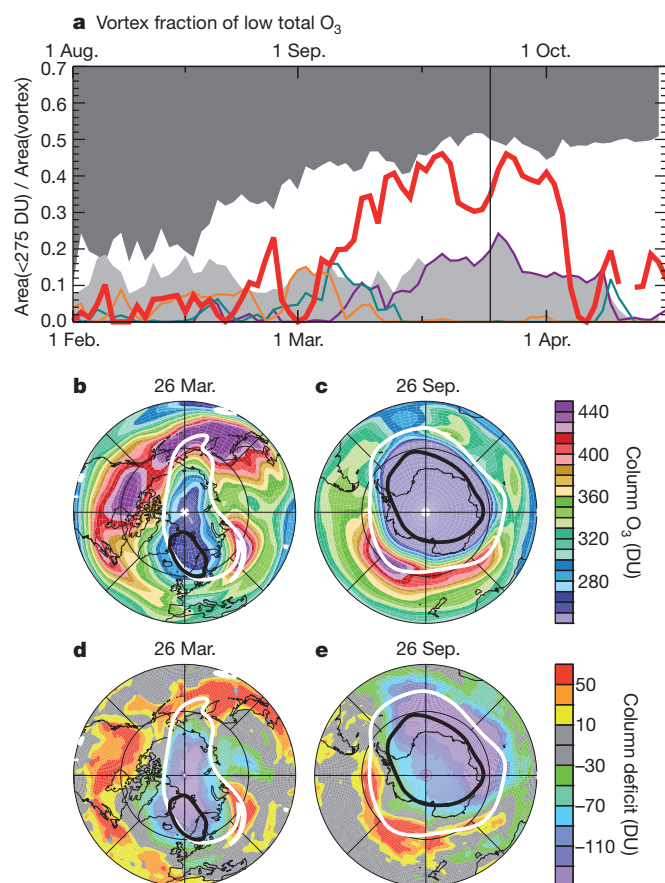


Figure 5 | Total column ozone. **a**, Time series of the fraction of 460 K vortex area with total ozone below 275 Dobson units (DU) in February–April in the Arctic (bottom axis), and in August–October in the Antarctic (top axis). Line colours/shading as in Fig. 1. 2005–2011 values are from OMI; earlier values are from TOMS (Total Ozone Mapping Spectrometer) instruments⁵⁰. Maps show OMI total ozone (**b**, **c**) and ozone deficit (**d**, **e**) in the Arctic (Antarctic) on 26 March 2011 (26 September 2010). Overlays as in Fig. 2 but at 460 K.

An echo of the Antarctic

In the absence of chemical ozone loss, downward transport during winter results in a springtime maximum in total ozone; because this transport is stronger in the Arctic, background ozone levels there are ~100 DU higher than those in the Antarctic^{7,23}. Therefore Arctic spring total ozone could, even after chemical destruction comparable to that in an Antarctic ozone hole (commonly defined by values less than 220 DU; refs 7, 12), exhibit only a weak maximum in total ozone rather than a well-defined minimum. Examination of the long-term ozone-sonde record in the Arctic shows that abundances near 250 DU or less are well below typical autumn values, thus appearing as a 'hole' in total ozone. Dynamical processes can result in transient regions of very low total ozone (Supplementary Discussion, Supplementary Figs 5, 6) and/or local minima in lower-stratospheric ozone profiles (for example, via ozone-poor extra-vortex air transported into the polar vortex^{14,24}). For an interhemispheric comparison of chemical loss, it is thus important to verify that observed Arctic ozone decreases were primarily related to chemical, rather than dynamical, processes.

Figure 4 shows that the precipitous decline in Arctic ozone in February–March 2011 resulted from chemical loss of similar magnitude to that in the Antarctic in the mid-1980s. Observed ozone between ~15 and 20 km altitude decreased to values matching the minima in early Antarctic ozone holes and those reached at the corresponding time in some recent Antarctic winters (Figs 2j–l; 3f). In late March–early April, most ozone-sonde profiles in the vortex had mixing ratios less than 1 p.p.m.v., with values ~0.7 p.p.m.v. over an approximately 2-km altitude region, and some dipping to 0.5 p.p.m.v. (Supplementary Fig. 7). Minimum total ozone in spring 2011 was continuously below 250 DU for ~27 days (Supplementary Fig. 5b), with a maximal area below that level of ~2 × 10⁶ km² (roughly five times the area of Germany or California). Values dropped to ~220–230 DU for about a week in late March 2011.

In these respects, chemical ozone destruction in the 2011 Arctic polar vortex attained, for the first time, a level clearly identifiable as an Arctic ozone hole. On the other hand, although the magnitude of chemical depletion was comparable to that in the Antarctic, total ozone values remained higher and, because the areal extent of the Arctic vortex was much smaller (~60% the size of a typical Antarctic vortex), the low-ozone region was more confined.

The Arctic winter stratosphere exhibits striking interannual variability. The past decade has included the four most dynamically active (hence among the warmest) Arctic winters in the past 32 years (ref. 35) and now the two coldest winters with largest ozone loss^{7,12–14}, extending the previously noted trend of the coldest winters becoming colder^{13,16}. Had implementation of the Montreal Protocol not curbed the increase in stratospheric halogen loading, formation of an Arctic ozone hole would have already become common even in moderately cold winters³⁶. Even with the lower anthropogenic halogen levels actually reached, the potential for Antarctic-like ozone loss in the Arctic in the event of a persistently cold winter–spring such as that in 2010–11 has been recognized for decades^{5,22}. Despite temperatures that were generally far higher than those in Antarctic winter, Arctic chemical ozone destruction in 2011 rivalled that in some Antarctic ozone holes. The development of an Arctic ozone hole under conditions only slightly more extreme than those in some previous Arctic winters raises the possibility of yet more severe depletion as lower-stratospheric temperatures decrease. More acute Arctic ozone destruction could exacerbate biological risks from increased ultraviolet radiation exposure, especially if the vortex shifted over densely populated mid-latitudes, as it did in April 2011.

Our present understanding of what drives variability in the Arctic winter stratosphere is incomplete. Stratospheric temperatures and vortex evolution depend on the atmosphere's radiative properties and propagation of wave activity^{37,38}, which are being modified by increasing greenhouse gas concentrations. Day-to-day tropospheric disturbances can lead to stratospheric warming or cooling, depending

on their geographical location and the stratospheric vortex structure, which controls their upward propagation^{39,40}. Current climate models do not fully capture either the observed short-timescale patterns of Arctic variability or the full extent of the observed longer-term cooling trend in cold stratospheric winters; nor do they agree on future circulation changes that affect trends in transport^{41,42}. Our ability to predict when conditions similar to, or more extreme than, those in 2011 may be realized is thus very limited. Improving our predictive capabilities for Arctic ozone loss, especially while anthropogenic halogen levels remain high, is one of the greatest challenges in polar ozone research. Comprehensive stratospheric data sets, such as those used here, are critical to meeting that challenge.

METHODS SUMMARY

MERRA (Modern Era Retrospective-analysis for Research and Applications⁴³) fields are used for temperature and vortex analysis and for vortex averaging of composition measurements. The CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) on the CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) satellite⁴⁴ provides PSC/aerosol information.

Trace gas profiles are from the Microwave Limb Sounder (MLS)⁴⁵ on NASA's Aura satellite. Only daytime CIO measurements are used. Northern (southern) high latitudes are sampled near midday (in late afternoon), thus the average solar zenith angle (SZA) of MLS Antarctic measurements is $\sim 7^\circ$ higher than that in the Arctic. Reactive chlorine partitioning shifts away from CIO at higher SZAs^{7,12}, leading to $\sim 30\%$ lower CIO measured in the Antarctic than in the Arctic under fully activated conditions. An instrument anomaly disrupted MLS measurements from 27 March to 20 April 2011. UARS (Upper Atmosphere Research Satellite) MLS measurements, used for 1995–1996 and 1996–1997 analyses, are sparse because of the UARS yaw cycle and other measurement gaps⁴⁶.

Total column ozone is measured by the Dutch-Finnish Ozone Monitoring Instrument (OMI)⁴⁶ on Aura. Total ozone 'deficit' is the difference between daily values and a reference that is minimally affected by chemical loss.

Measurements from MLS and the Match network of balloon-borne ozone soundings (ozone sondes)⁴⁷ are used to estimate chemical ozone loss in two ways. The difference between calculated 'passive' (influenced only by transport) ozone and observed ozone is computed, with passive ozone obtained using MLS nitrous oxide¹⁴, a 'reverse trajectory' model^{25,26}, and the ATLAS (Alfred Wegener Institute Lagrangian Chemistry/Transport System) model²⁷. Vortex ozone is also examined on the surfaces on which it subsides^{12,14,28,48}, with descent rates from modelled radiative heating/cooling rates averaged over the polar vortex⁴⁸.

Photochemical box model runs were performed using the chemical model from ATLAS²⁷ to test the sensitivity of ozone loss to initial ozone amounts and denitrification.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 May; accepted 7 September 2011.

Published online 2 October 2011.

- Farman, J. C., Gardiner, B. G. & Shanklin, J. D. Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature* **315**, 207–210 (1985).
- Solomon, S., Garcia, R. R., Rowland, F. S. & Wuebbles, D. J. On the depletion of Antarctic ozone. *Nature* **321**, 755–758 (1986).
- Molina, L. T. & Molina, M. J. Production of Cl_2O_2 from the self-reaction of the CIO radical. *J. Phys. Chem.* **91**, 433–436 (1987).
- Anderson, J. G., Brune, W. H. & Proffitt, M. H. Ozone destruction by chlorine radicals within the Antarctic vortex: the spatial and temporal evolution of CIO- O_3 anticorrelation based on *in situ* ER-2 data. *J. Geophys. Res.* **94**, 11465–11479 (1989).
- Solomon, S. Stratospheric ozone depletion: a review of concepts and history. *Rev. Geophys.* **37**, 275–316 (1999).
- Schoeberl, M. R. & Hartmann, D. L. The dynamics of the stratospheric polar vortex and its relation to springtime ozone depletions. *Science* **251**, 46–52 (1991).
- World Meteorological Organization. *Scientific Assessment of Ozone Depletion: 2010* (Report 52, Global Ozone Research and Monitoring Project, 2011).
- Solomon, P. M. *et al.* High concentrations of chlorine monoxide at low altitudes in the Antarctic spring stratosphere: secular variation. *Nature* **328**, 411–413 (1987).
- Waters, J. W. *et al.* Stratospheric CIO and ozone from the Microwave Limb Sounder on the Upper Atmosphere Research Satellite. *Nature* **362**, 597–602 (1993).
- Santee, M. L., Manney, G. L., Waters, J. W. & Livesey, N. J. Variations and climatology of CIO in the polar lower stratosphere from UARS Microwave Limb Sounder measurements. *J. Geophys. Res.* **108**, 4454, <http://dx.doi.org/10.1029/2002JD003335> (2003).

- Santee, M. L. *et al.* A study of stratospheric chlorine partitioning based on new satellite measurements and modeling. *J. Geophys. Res.* **113**, D12307, <http://dx.doi.org/10.1029/2007JD009057> (2008).
- World Meteorological Organization. *Scientific Assessment of Ozone Depletion: 2006* (Report 50, Global Ozone Research and Monitoring Project, 2007).
- Rex, M. *et al.* Arctic winter 2005: implications for stratospheric ozone loss and climate change. *Geophys. Res. Lett.* **33**, L23808, <http://dx.doi.org/10.1029/2006GL026731> (2006).
- Manney, G. L. *et al.* EOS MLS observations of ozone loss in the 2004–2005 Arctic winter. *Geophys. Res. Lett.* **33**, L04802, <http://dx.doi.org/10.1029/2005GL024494> (2006).
- Harris, N. R. P., Lehmann, R., Rex, M. & von der Gathen, P. A closer look at Arctic ozone loss and polar stratospheric clouds. *Atmos. Chem. Phys.* **10**, 8499–8510 (2010).
- Rex, M. *et al.* Arctic ozone loss and climate change. *Geophys. Res. Lett.* **31**, L04116, <http://dx.doi.org/10.1029/2003GL018844> (2004).
- Tilmes, S., Müller, R., Engel, A., Rex, M. & Russell, J. M. III. Chemical ozone loss in the Arctic and Antarctic stratosphere between 1992 and 2005. *Geophys. Res. Lett.* **33**, L20812, <http://dx.doi.org/10.1029/2006GL026925> (2006).
- Poole, L. R. & Pitts, M. C. Polar stratospheric cloud climatology based on Stratospheric Aerosol Measurement II observations from 1978 to 1989. *J. Geophys. Res.* **99**, 13083–13089 (1994).
- Fromm, M. D. *et al.* An analysis of Polar Ozone and Aerosol Measurement (POAM) II Arctic stratospheric cloud observations, 1993–1996. *J. Geophys. Res.* **104**, 24341–24357 (1999).
- Pitts, M. C., Poole, L. R. & Thomason, L. W. CALIPSO polar stratospheric cloud observations: second-generation detection algorithm and composition discrimination. *Atmos. Chem. Phys.* **9**, 7577–7589 (2009).
- Douglass, A. R. *et al.* Interhemispheric differences in springtime production of HCl and ClONO_2 in the polar vortices. *J. Geophys. Res.* **100**, 13967–13978 (1995).
- Manney, G. L. *et al.* Chemical depletion of ozone in the Arctic lower stratosphere during winter 1992–93. *Nature* **370**, 429–434 (1994).
- Tegtmeier, S., Rex, M., Wohltmann, I. & Krüger, K. Relative importance of dynamical and chemical contributions to Arctic wintertime ozone. *Geophys. Res. Lett.* **35**, L17801, <http://dx.doi.org/10.1029/2008GL034250> (2008).
- Rex, M. *et al.* *In situ* measurements of stratospheric ozone depletion rates in the Arctic winter of 1991/1992: a Lagrangian approach. *J. Geophys. Res.* **103**, 5843–5853 (1998).
- Manney, G. L. *et al.* Lagrangian transport calculations using UARS data. Part II: ozone. *J. Atmos. Sci.* **52**, 3069–3081 (1995).
- Manney, G. L. *et al.* Variability of ozone loss during Arctic winter (1991–2000) estimated from UARS Microwave Limb Sounder measurements. *J. Geophys. Res.* **108**, 4149, <http://dx.doi.org/10.1029/2002JD002634> (2003).
- Wohltmann, I., Lehmann, R. & Rex, M. The Lagrangian chemistry and transport model ATLAS: simulation and validation of stratospheric chemistry and ozone loss in the winter 1999/2000. *Geosci. Model Dev.* **3**, 585–601 (2010).
- von der Gathen, P. *et al.* Observational evidence for chemical ozone depletion over the Arctic in winter 1991–92. *Nature* **375**, 131–134 (1995).
- Gernandt, H. The vertical ozone distribution above the GDR-research base, Antarctica in 1985. *Geophys. Res. Lett.* **14**, 84–86 (1987).
- Rex, M. *et al.* Prolonged stratospheric ozone loss in the 1995–96 Arctic winter. *Nature* **389**, 835–838 (1997).
- Manney, G. L., Froidevaux, L., Santee, M. L., Zurek, R. W. & Waters, J. W. MLS observations of Arctic ozone loss in 1996–97. *Geophys. Res. Lett.* **24**, 2697–2700 (1997).
- Manney, G. L., Santee, M. L., Froidevaux, L., Waters, J. W. & Zurek, R. W. Polar vortex conditions during the 1995–96 Arctic winter: meteorology and MLS ozone. *Geophys. Res. Lett.* **23**, 3203–3206 (1996).
- Petzoldt, K. The role of dynamics in total ozone deviations from their long-term mean over the Northern Hemisphere. *Ann. Geophys.* **17**, 231–241 (1999).
- Hood, L. L., Soukharev, B. E., Fromm, M. & McCormack, J. P. Origin of extreme ozone minima at middle to high northern latitudes. *J. Geophys. Res.* **106**, 20925–20940 (2001).
- Manney, G. L. *et al.* Aura Microwave Limb Sounder observations of dynamics and transport during the record-breaking 2009 Arctic stratospheric major warming. *Geophys. Res. Lett.* **36**, L12815, <http://dx.doi.org/10.1029/2009GL038586> (2009).
- Newman, P. A. *et al.* What would have happened to the ozone layer if chlorofluorocarbons (CFCs) had not been regulated? *Atmos. Chem. Phys.* **9**, 2113–2128 (2009).
- Newman, P. A., Nash, E. R. & Rosenfield, J. E. What controls the temperatures of the Arctic stratosphere during the spring? *J. Geophys. Res.* **106**, 19999–20010 (2001).
- Polvani, L. M. & Saravanan, R. The three-dimensional structure of breaking Rossby waves in the polar wintertime stratosphere. *J. Atmos. Sci.* **57**, 3663–3685 (2000).
- Orsolini, Y. J., Karpechko, A. Y. & Nikulin, G. Variability of the Northern Hemisphere polar stratospheric cloud potential: the role of North Pacific disturbances. *Q. J. R. Meteorol. Soc.* **135**, 1020–1029 (2009).
- Woollings, T., Charlton-Perez, A., Ineson, S., Marshall, G. & Masato, G. Associations between stratospheric variability and tropospheric blocking. *J. Geophys. Res.* **115**, D06108, <http://dx.doi.org/10.1029/2009JD012742> (2010).
- Butchart, N. *et al.* Multimodel climate and variability of the stratosphere. *J. Geophys. Res.* **116**, D05102, <http://dx.doi.org/10.1029/2010JD014995> (2011).
- Charlton-Perez, A. *et al.* The potential to narrow uncertainty in projections of stratospheric ozone over the 21st century. *Atmos. Chem. Phys.* **10**, 9473–9486 (2010).
- Reinecker, M. M. *et al.* MERRA — NASA's modern-era retrospective analysis for research and applications. *J. Clim.* **24**, 3624–3648 <http://dx.doi.org/10.1175/JCLI-D-11-00015.1> (2011).

44. Hunt, W. H. *et al.* CALIPSO lidar description and performance assessment. *J. Atmos. Ocean. Technol.* **26**, 1214–1228 (2009).
45. Waters, J. W. *et al.* The Earth Observing System Microwave Limb Sounder (EOS MLS) on the Aura satellite. *IEEE Trans. Geosci. Rem. Sens.* **44**, 1075–1092 (2006).
46. Levelt, P. F. *et al.* The Ozone Monitoring Instrument. *IEEE Trans. Geosci. Rem. Sens.* **44**, 1093–1101 (2006).
47. Rex, M. *et al.* Chemical ozone loss in the Arctic winter 1994/95 as determined by the Match technique. *J. Atmos. Chem.* **32**, 35–59 (1999).
48. Rex, M. *et al.* Chemical depletion of Arctic ozone in winter 1999/2000. *J. Geophys. Res.* **107**, 8276, <http://dx.doi.org/10.1029/2001JD000533> (2002).
49. Hoskins, B. J., McIntyre, M. E. & Robertson, A. W. On the use and significance of isentropic potential-vorticity maps. *Q. J. R. Meteorol. Soc.* **111**, 877–946 (1985).
50. McPeters, R. D. *et al.* *Earth Probe Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide* (NASA Technical Publication 1998-206895, 1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the MLS (especially A. Lambert, D. Miller, W. Read, M. Schwartz, P. Stek, J. Waters), OMI (especially P. K. Bhartia, G. Jaross, G. Labow), CALIPSO and Match science teams, as well as A. Douglass, J. Joiner and the Aura project, for their support. We also thank W. Daffer and R. Fuller for programming assistance at JPL; the many observers whose work went into obtaining the ozone-sonde measurements; the ozone scientists who participated in the discussion of the 2011 Arctic ozone loss and appropriate definition of an Arctic ozone hole (including, but not limited to, N. Harris, G. Bodeker, G. Braathen, M. Kurylo, R. Salawitch); and especially P. Newman and K. Minschwaner for discussions and comments. Meteorological analyses were provided by NASA's Global Modeling and Assimilation Office (GMAO) and by the European Centre for Medium-Range Weather Forecasts. We thank S. Pawson of GMAO for advice on usage of the MERRA reanalysis. Ozone-sonde measurements at Alert, Eureka, Resolute Bay, Churchill and Goose Bay were funded by Environment Canada. Additional ozone sondes were flown at Eureka as part of the

Canadian Arctic Atmospheric Chemistry Experiment (ACE) Validation Campaign and were funded by the Canadian Space Agency. Academy of Finland provided partial funding for performing and processing ozone-sonde measurements in Jokioinen and Sodankylä. Ozone soundings and work at AWI were partially funded by the EC DG Research through the RECONCILE project. Work at the Jet Propulsion Laboratory, California Institute of Technology, and at Science Systems and Applications Inc., was done under contract with NASA.

Author Contributions G.L.M. and M.L.S. led analysis of MLS data; M.R. led analysis of ozone-sonde data; G.L.M. led the meteorological data analysis. M.R., G.L.M., N.J.L. and I.W. did chemical ozone loss calculations. R.L. and M.R. performed and analysed chemical box model calculations. M.C.P. and L.R.P. provided CALIPSO/CALIP data analyses; E.R.N. and P.V. provided TOMS and OMI data analyses. L.F., M.L.S., G.L.M. and N.J.L. provided expertise on MLS data usage; D.P.H., P.V. and P.F.L. provided expertise on OMI data usage. J.D., V.D., H.G., B.J., R.K., E.K., N.L., A.M., C.T.M., H.N., M.C.P., D.W.T., P.v.d.G., K.A.W. and N.S.Z. were responsible for performing and processing ozone-sonde measurements. All authors contributed comments on the manuscript. G.L.M., M.L.S. and M.R. jointly compiled and synthesized the results. G.L.M. and M.L.S. wrote the paper.

Author Information CALIP data are publicly available at http://eosweb.larc.nasa.gov/PRODOCS/calipso/table_calipso.html, MLS data at <http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/MLS>, OMI data at http://disc.sci.gsfc.nasa.gov/Aura/data-holdings/OMI/omto3_v003.shtml, and GEOS-5 MERRA analyses through <http://disc.sci.gsfc.nasa.gov/mdisc/data-holdings/merra/>. The balloon-borne Antarctic ozone-sonde data recorded in 1985 and the following years are publicly available at <http://dx.doi.org/10.1594/PANGAEA.547983>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.L.M. (Gloria.L.Manney@jpl.nasa.gov) or M.L.S. (Michelle.L.Santee@jpl.nasa.gov).

METHODS

Data sets. Modern Era Retrospective-analysis for Research and Applications (MERRA)⁴³ fields, from the Goddard Earth Observing System Version 5.2.0 (GEOS-5) data assimilation system, are used for the temperature and vortex analysis. The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) on the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) satellite⁴⁴ provides PSC/aerosol information. CALIOP measurements began in April 2006. Trace gas profile measurements are from the Microwave Limb Sounder (MLS)⁴⁵ on NASA's Aura satellite, and the predecessor MLS instrument²⁶ on the Upper Atmosphere Research Satellite (UARS). Total column ozone data are from the Dutch-Finnish Ozone Monitoring Instrument (OMI)⁴⁶ on board Aura. The historical total ozone record comprises data from Nimbus-7 and Earth Probe Total Ozone Mapping Spectrometer (TOMS)⁵⁰. Aura MLS and OMI measurements are available from August 2004 through to the present. UARS MLS measurements were obtained from September 1992 through to early 2000, with increasingly sparse sampling in the later years²⁶. TOMS data are available beginning in 1979, but no TOMS instrument was taking measurements during the 1995–96 Arctic winter.

Measurements from the Match network of balloon-borne ozone soundings (ozone sondes)⁴⁷ are used in some of the chemical ozone loss estimates.

Temperature and vortex analysis. Potential vorticity⁴⁹ (PV) is used to define the vortex, with a contour of 'scaled' PV of $1.4 \times 10^{-4} \text{ s}^{-1}$ (in vorticity units) demarking the vortex edge^{51,52}. Vortex strength is diagnosed as the maximum daily gradient in PV as a function of equivalent latitude (the latitude that would enclose the same area between it and the pole as a given PV contour)^{51–53}. Scaled PV multiplied by 10^4 is used in the calculation, resulting in units for its gradient of $10^{-4} \text{ (s degrees equivalent latitude)}^{-1}$.

The temperature threshold for chlorine activation, T_{act} , is estimated using the formula for nitric acid trihydrate formation⁵⁴, which depends on pressure, HNO_3 and H_2O . Climatological HNO_3 and H_2O profiles are used, derived from UARS data. The area with $T < T_{\text{act}}$ is calculated on seven isentropic surfaces in the lower stratosphere: 390, 410, 430, 460, 490, 520 and 550 K; T_{act} on these levels is 197.5, 197.2, 196.8, 196.5, 195.9, 195.3 and 194.5 K, respectively. To get the volume with $T < T_{\text{act}}$ from 380 through 565 K, the areas at each of the seven levels are multiplied by the estimated altitude associated with that layer and summed. The altitude range associated with each layer is obtained from a standard potential temperature profile as a function of altitude derived from high latitude temperature soundings taken during the 1988–89 through to 2001–02 winters (the same profile was used for V_{psc} calculations in refs 13, 16 and 48). These thicknesses are 1.29088, 1.19995, 1.36770, 1.46281, 1.30554, 1.18199 and 1.07382 km for the seven levels listed above. Vortex volume is calculated from vortex area in the same manner. Winter mean V_{psc} is calculated over 16 December through to 15 April. Previous studies have shown that V_{psc} scaled by the vortex area is a good proxy for chlorine activation and ozone loss potential¹⁷. Additional temperature and vortex diagnostics are described in Supplementary Information.

Polar stratospheric cloud and aerosol information. Particulate backscatter averaged over the polar vortex derived from CALIOP data is used to provide PSC/aerosol information. Total attenuated backscatter at 532 nm, $b(z)$, is one of the basic CALIOP Level 1B data products. $b(z)$ is the sum of the particulate backscatter (due to liquid aerosol and PSCs), $b_p(z)$, and molecular backscatter, $b_m(z)$. $b_m(z)$ is calculated using GEOS-5 molecular density profiles (included in the CALIOP Level 1B data files) and a theoretical value for the molecular scattering cross-section⁵⁵. Profiles of $b_p(z)$ are then produced by subtracting $b_m(z)$ from $b(z)$. Vortex-averaged profiles of $b_p(z)$ are produced by averaging all CALIOP $b_p(z)$ profiles located inside the vortex edge (defined using information available in GEOS-5 Derived Meteorological Product (DMP) files for the nearly-coincident Aura MLS data⁵²) over the selected time interval.

MLS trace gas profile measurements and analysis. Trace gas profile measurements of HNO_3 , HCl, ClO, ozone and N_2O (a long-lived tracer used to assess descent) are from Aura MLS⁴⁵ version 3 retrievals; data quality screening is as recommended in the MLS data quality document⁵⁶. MLS data are retrieved on pressure surfaces; potential temperature as a function of pressure from MLS DMPs⁵² calculated from GEOS-5 analyses is used to interpolate to isentropic surfaces. Vortex averages of MLS data are calculated using the $1.4 \times 10^{-4} \text{ s}^{-1}$ scaled PV contour to define the vortex edge, using PV values from the MLS DMPs⁵². Active chlorine is in the form of ClO mainly during the daytime, and thus measured ClO amounts vary with the solar zenith angle (SZA) at which the measurements are taken. Only daytime ClO measurements are used here. Northern high latitudes are sampled near midday local time, southern high latitudes are sampled in late afternoon, thus the SZA of Aura MLS Antarctic measurements is $\sim 7^\circ$ higher on average than that in the Arctic. Reactive chlorine partitioning shifts away from ClO at higher SZAs^{7,12}, leading to $\sim 30\%$ lower ClO measured by Aura MLS in the Antarctic than in the Arctic under fully activated

conditions. MLS measurements are unavailable from 27 March through to 20 April 2011 because of an instrument anomaly. Upper Atmosphere Research Satellite (UARS) MLS measurements, used for analysis of 1995–96 and 1996–97, are sparse because of the UARS yaw cycle and other measurement gaps²⁶. The time of day of UARS measurements varied through the yaw cycle, in the middle of which no daytime ClO measurements were obtained¹⁰; thus ClO values shown in 1995–96 and 1996–97 near those dates (including the mid-February 1996 measurements shown in Fig. 2g) are not representative of the degree of chlorine activation.

Chemical loss calculations. Chemical ozone loss is quantified by two methods, both widely used for such calculations^{7,12,24–28,47,48}. In the 'passive subtraction' method^{25–27}, a transport model is used to calculate the evolution of ozone in the absence of chemical changes ('passive' ozone). The difference between passive ozone and observed ozone provides an estimate of chemical loss.

Here, passive ozone is obtained in three different ways. First, MLS observations of N_2O , a long-lived species unaffected by chemical processes, are used to calculate vertical motion, and that estimate of descent is then used to calculate how initial MLS ozone profiles would have evolved in the absence of chemical loss¹⁴. Second, a 'reverse trajectory' transport model^{25,26} is used to transport an initial state based on MLS-observed ozone with no chemistry. Finally, the ATLAS (Alfred Wegener Institute Lagrangian Chemistry/Transport System) chemistry and transport model is run in passive mode²⁸, initialized with MLS ozone.

Vortex ozone is also examined in relation to the surfaces on which it is subsiding^{12,14,28,48}. The descent rates used here are obtained by averaging radiative heating/cooling rates from the radiation calculation used in the ATLAS model over the polar vortex⁴⁸. These rates are then used to examine vortex-averaged MLS and ozone-sonde data on surfaces of 'spring equivalent potential temperature'⁴⁸, defined as the potential temperature at which air originating at a given level arrived at the end of March. Since the air descended on these surfaces, ozone would have been constant on each such surface in the absence of chemical loss.

The ozone-sonde data used here are all from electrochemical concentration cell (ECC) sondes, made by different manufacturers. Ozone-sonde data quality was assessed in an intercomparison experiment⁵⁷ and is discussed in ref. 47. For chemical loss calculations using ozone-sonde data, the profiles are first examined using a procedure for detecting lamination in the profiles; such lamination (an example is shown in Fig. 3f) is associated with mixing in of extra-vortex air, which may obscure the signature of chemical loss. Profiles that have been significantly altered by mixing processes, as indicated by lamination, are excluded from the vortex averages used in the chemical loss calculations. 2010–11 Arctic ozone-sonde data are provided as Supplementary Information.

Results from the ATLAS model passive subtraction calculations, and from the calculations on spring equivalent potential temperature surfaces using the Match network ozone-sonde data, are shown in Fig. 4; all panels show vortex averages. These results have been compared with the results from the other methods described above. While absolute ozone values obtained from different methods/data sets vary significantly (up to $\sim 0.4 \text{ p.p.m.v.}$ at the end of March 2011), the year-to-year variations in chemical loss calculated using all three methods agree closely, indicating a high degree of precision in the relative amount of calculated loss between different years.

The Alfred Wegener Institute chemical box model, also used as the chemical module in ATLAS, simulates 175 reactions between 48 chemical species in the stratosphere^{27,58}. This model was used to perform conceptual runs (Supplementary Fig. 4), started on 1 March with identical initial mixing ratios of all species except HNO_3 and O_3 . For these two species values corresponding to 1997 (3 p.p.m.v. O_3 , 10 p.p.b.v. HNO_3) and 2011 (2.2 p.p.m.v. O_3 , 6 p.p.b.v. HNO_3) (compare Figs 2a and 4c) were combined to yield four sets of initial conditions. Initial ClO_x was 2 p.p.b.v., corresponding to the vortex-averaged ClO_x derived by ATLAS from MLS ClO measurements on 1 March 2011. An air parcel at 70°N , 460 K potential temperature, with a temperature of 193 K throughout March, was used. Heterogeneous reactions took place on liquid aerosols, rather than solid (nitric acid trihydrate, NAT) PSCs, since the widespread existence of the latter is inconsistent with MLS observations of gas-phase HNO_3 values (Fig. 2a) larger than those the microphysical module predicts if NAT is present. A sensitivity run showed that sporadically occurring solid PSCs did not change the results significantly.

Column ozone and ozone deficit calculation. OMI total ozone data were processed with version 8.5 of the TOMS algorithm and have been extensively validated⁵⁹. TOMS data were processed with version 8 of the algorithm. The OMI and TOMS total ozone data used in this study were averaged on a fixed global $1^\circ \times 1^\circ$ latitude \times longitude grid. Averages were computed by area-weighting observations based on the overlap of their instantaneous field-of-view with each grid cell. Only data that satisfy quality criteria based on measurement path length and algorithm diagnostic criteria were included in the averaged samples.

Individual total ozone retrievals included in the samples are expected to have a root-mean-squared error of 1–2%.

Total ozone 'deficit' is calculated as the difference between daily values and a reference that is minimally affected by chemical ozone loss. The reference for the Arctic is the daily mean over all Arctic winters from 1978–79 through to 2009–10, from OMI starting in 2004–05 and from TOMS for earlier years⁵⁰. The Antarctic reference state is the daily mean of TOMS measurements for 1979 through to 1981. Because the Antarctic reference state is based on only three years' data for each day, variations in vortex position are not effectively averaged out; this reference is thus less robust than that for the Arctic, so patterns in daily maps may partially reflect differences in vortex position between the reference and the focus day.

51. Manney, G. L., Zurek, R. W., Gelman, M. E., Miller, A. J. & Nagatani, R. The anomalous Arctic lower stratospheric polar vortex of 1992–1993. *Geophys. Res. Lett.* **21**, 2405–2408 (1994).
52. Manney, G. L. *et al.* Solar occultation satellite data and derived meteorological products: Sampling issues and comparisons with Aura MLS. *J. Geophys. Res.* **112**, D24S50, <http://dx.doi.org/10.1029/2007JD008709> (2007).
53. Butchart, N. & Remsberg, E. E. The area of the stratospheric polar vortex as a diagnostic for tracer transport on an isentropic surface. *J. Atmos. Sci.* **43**, 1319–1339 (1986).
54. Hanson, D. & Mauersberger, K. Laboratory studies of the nitric acid trihydrate: implications for the south polar stratosphere. *Geophys. Res. Lett.* **15**, 855–858 (1988).
55. Hostetler, C. A. *et al.* CALIOP algorithm theoretical basis document. Calibration and Level 1 data products (Technical Report, NASA Langley Research Center, 2006); available at (<http://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-201v1.0.pdf>).
56. Livesey, N. J. *et al.* Version 3.3 Level 2 data quality and description document. (Technical Report JPL D-33509, Jet Propulsion Laboratory, 2010); available at (http://mls.jpl.nasa.gov/data/v3-3_data_quality_document.pdf).
57. Smit, H. G. *et al.* Assessment of the performance of ECC-ozone sondes under quasi-flight conditions in the environmental simulation chamber: insights from the Jülich Ozone Sonde Intercomparison Experiment (JOSIE). *J. Geophys. Res.* **112**, D19306, <http://dx.doi.org/10.1029/2006JD007308> (2007).
58. Krämer, M. *et al.* Intercomparison of stratospheric chemistry models under polar vortex conditions. *J. Atmos. Chem.* **45**, 51–77 (2003).
59. McPeters, R. *et al.* Validation of the Aura Ozone Monitoring Instrument total column ozone product. *J. Geophys. Res.* **113**, D15S14, <http://dx.doi.org/10.1029/2007JD008802> (2008).